5-31-2023

EXPLORING SENIOR CHEMICAL ENGINEERING STUDENTS' APPROACH TO PROCESS SAFETY DECISION MAKING

Jeffrey Saunders Stransky
Rowan University

EXPLORING SENIOR CHEMICAL ENGINEERING STUDENTS' APPROACH TO PROCESS SAFETY DECISION MAKING

by
Jeffrey Saunders Stransky

A Dissertation

Submitted to the
Department of Experiential Engineering Education
College of Engineering
In partial fulfillment of the requirement
For the degree
Doctor of Philosophy
at
Rowan University
March 31, 2023

Dissertation Chair: Cheryl Bodnar, Ph.D., Associate Professor of Experiential Engineering Education

Committee Members:
Juan Cruz, Ph.D., Assistant Professor, Department of Experiential Engineering Education, Rowan University
Kevin Dahm, Ph.D., Professor, Department of Chemical Engineering, Rowan University
Emily Dringenberg, Ph.D., Assistant Professor, Department of Engineering Education, The Ohio State University

Acknowledgments

To Dr. Cheryl Bodnar, thank you for always having faith in me. Your diligent feedback and mentorship molded me into the researcher who could write this dissertation. I cannot express how much I appreciate your investment in me.

To Dr. Juan Cruz, Dr. Kevin Dahm, and Dr. Emily Dringenberg, thank you for your input and expertise through many conversations. Your guidance has been invaluable.

To my colleagues in ExEEd, new and old, I could not have asked for a better home on my journey. Thank you for challenging and encouraging me along the way.

To Trent and Ian, thank you for your bond of friendship and the many hours we've spent together. The late weeknights we shared breathed life and laughter back into my week.

To Mom and Dad, thank you for every opportunity you have given me to send me on my journey. I could not have asked for more loving and supporting parents.

To my wonderful partner, Melanie, thank you for supporting me and my interest in science as you have since day one. You add so much color to my life, and I would not have completed this material without your never-ending encouragement. I love you.

This material has been supported by the National Science Foundation Improving Undergraduate STEM Education [IUSE DUE#1711376, 1711644, 1711672, and 1711866] and the US Department of Education Graduate Assistance in Areas of National Need Fellowship program [P200A180055]. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of these entities.

Abstract

Jeffrey Stransky
EXPLORING SENIOR CHEMICAL ENGINEERING STUDENTS' APPROACH TO PROCESS SAFETY DECISION MAKING
2022-2023
Cheryl Bodnar, Ph.D.
Doctor of Philosophy

Despite investments in chemical process safety education, evidence suggests that engineers' decision making may contribute to process safety incidents. Currently, limited educational endeavors in process safety decision making exist, raising the need for a better understanding of how to prepare chemical engineering students for industry decisions. This dissertation fortifies current process safety education through three studies involving senior chemical engineering students. Study One developed the Engineering Process Safety Research Instrument (EPSRI) through exploratory and confirmatory factor analyses. Educators may use the EPSRI to evaluate their students' moral development. Study Two evaluated a digital process safety game, *Contents Under Pressure*, on students' moral decision making. Study Two supports supplementing process safety curriculum with digital games because the game simulates process safety decisions in a plant environment, providing immersive first-hand experience without real-world risks. Study Two showed that students who played Contents Under Pressure made authentic decisions in the game, advancing their moral development. Study Three explored how students consider process safety criteria when making decisions, finding that safety takes precedence over other criteria, such as productivity, despite potential negative implications. Study Three supports educators contextualizing incident case studies with discussion on how process safety criteria compete.

Table of Contents

Table of Contents (Continued)

Table of Contents (Continued)

Table of Contents (Continued)

Table of Contents (Continued)

List of Figures

List of Figures (Continued)

Figure Page

List of Tables

Table Page

List of Tables (Continued)

Table Page

List of Tables (Continued)

Table Page

Chapter 1

Introduction

Background

The chemical process industry has increasingly become concerned with the development of process safety over the past 40 years. These efforts have been manifested through the development of AIChE's Center for Chemical Process Safety (CCPS) and Safety and Chemical Engineering Education (SAChE) programs (CCPS, 2022; Safety and Chemical Engineering Education, 2022; Willey, 1999), the formation of the US Chemical Safety and Hazard Investigation Board (CSB) (2009) and European Chemical Agency (ECHA) (2023), and the inclusion of process safety as an educational outcome from the Accreditation Board for Engineering and Technology (ABET) (ABET, 2021a; Vaughen, 2012). Efforts have primarily focused on integrating prevention through design (PtD) strategies (Biddle & Afanuh, 2015) and educating engineers in hazard and operability studies (HAZOP) and Layers of Protection Analysis (LOPA) (Cong et al., 2016; Willey et al., 2020).

However, the role of an engineer's decision making is frequently overlooked despite its critical impact on process safety and potential to inhibit preventative layers (CSB, 2017; Ness, 2015). Incident case studies performed by the CSB have frequently provided evidence that poor decision making can create pathways towards a process incident (CSB, 2011, 2015, 2018, 2019b). For instance, decision making has led to muting warning alarms, breaking protocols (CSB, 2019b), and rejecting a plant shut down during an active chemical leak (CSB, 2015), where each of these decisions caused or exacerbated a process incident that could have been avoided.

In such situations, engineers may not be intentionally devaluing process safety; instead, engineers may be overlooking the implications to their decision making. These behaviors may be driven by ethical biases (Bazerman & Tenbrunsel, 2011) or demands from competing criteria within each decision situation (Lehmann et al., 2022). When predicting their decisions to process safety situations, engineers may believe they will operate at an inflated level of ethicality based on what they think they *should* do, yet when confronted with those situations in practice, they may make decisions that are notably less ethical because the ethical dimensions of the situations may fade out of vision among other considerations (Bazerman & Tenbrunsel, 2011; Tenbrunsel & Messick, 2004). Moreover, safety goals may be overlooked as other criteria, such as production and budget, compete for priority within engineers' decision making forcing engineers to make trade-offs in criteria (Lehmann et al., 2022). SPHERA's annual safety report found that while safety goals are established, 69% of goals are not met due to competing criteria (Lehmann et al., 2022). Competing criteria are also prevalent in additional industries that rely on the decision making of practitioners (Akinleye et al., 2019; Dönmez & Uslu, 2020; Encinosa & Bernard, 2005; Hendrickson & Au, 2008; Lehmann et al., 2022).

While engineers' decision making has strong implications on process safety, engineers receive limited preparation throughout their education. For example, SAChE provides students with educational media, such as video case studies and worksheets (Safety and Chemical Engineering Education, 2022), but engineers still lack first hand practice making decisions that challenge their ethics or force them to make trade-offs in criteria. Firsthand practice is difficult to obtain particularly in process safety because of the hazards associated with the industry. Barab and Duffy (Barab & Duffy, 2012) call for "practice fields" where individuals may gain firsthand, practical experience without hazard. Game-based learning (GBL), such as through digital games, may supplement engineers' education by providing them with the opportunity to build this firsthand experience in decision making without the real world implications (Crawford, 1997). Albeit a safe alternative, GBL must be evaluated to understand its effectiveness as an educational tool for engineers' ethical decision making.

Study Objective

The objective of this work is to evaluate the effectiveness of a digital, process safety game as an educational intervention for engineering students' process safety decision making in contexts that challenge their ethics or force them to make trade-offs with competing criteria. The objective of this work is achieved through three studies. The first study sought to complete the development of the Engineering Process Safety Research Instrument (EPSRI) as a moral reasoning assessment in process safety contexts. The EPSRI provides a method to assess the effect of process safety curriculum, including digital GBL, on how engineers make ethical decisions relevant to process safety. The second study compared engineering students' moral reasoning related to process safety decisions within both a digital, process safety game (Contents Under Pressure) and traditional process safety curriculum. The third study investigated how engineering students approach decision making in process safety situations within the context of Contents Under Pressure where multiple criteria (budget, personal relationships plant productivity, safety, and time) compete for attention and priority. Not all process safety decisions have obvious ethical implications, so understanding how students approach process safety decisions with competing criteria provides a foundation for further pedagogical improvements in teaching process safety.

Study Scope

This work focuses on process safety decision making through the lens of moral reasoning and competing criteria. While there may be other approaches to evaluating engineering decision making, there is limited work on discrepancies between engineers' ethics and behavioral ethics. Thus, this study aims to provide a novel approach to mitigating process incidents caused by failures in engineering decision making by utilizing a digital, process safety game as a form of pedagogy. However, this study does not focus on the design of the game, its impact on student motivation, nor its assistance to students achieving course learning objectives. It does focus on evaluating its impact on ethical decision making through moral reasoning and on exploring how engineers treat criteria in process safety contexts. Last, this work only studies the decision making of senior chemical engineering students. The author purposefully sampled from this population because engineering educators most commonly incorporate process safety during this year of the curriculum (Mkpat et al., 2018). In the future, this work could be expanded to include industry practitioners during initial process safety training or review modules.

EPSRI instrument development in the first study entails continuing exploratory factor analysis (EFA) and completing a confirmatory factor analysis (CFA), which had begun in recent work (Butler, 2018). EFA and CFA essentially tests if the considerations designed into the EPSRI accurately and repeatedly represent levels of moral reasoning (Hair et al., 1998; Meyers et al., 2017a).

The second study leverages moral reasoning assessment from the EPSRI to compare Contents Under Pressure against traditional process safety curriculum. Comparisons are made based on (1) differences in moral reasoning between a comparison (traditional curriculum) and an intervention (Contents Under Pressure curriculum) cohort and (2) on differences between students' rated importance of considerations linked to different levels of moral reasoning in the EPSRI and in Contents Under Pressure. This work answers the following overarching research question, "*how does a digital, process safety game influence the ethical decision making of chemical engineering students?*"

The third study evaluates engineering students' process safety criteria priorities through both (1) a survey where they rank these criteria and then write an explanation for these rankings and (2) their play through of Contents Under Pressure where they rank these criteria inherently through their decision making. This work answers the following overarching research question, "*what role does competing criteria hold in the process safety decisions of engineering students?*"

Study Outcomes and Significance

Four significant outcomes were obtained as a result of the three studies. (1) **Development of the EPSRI was finalized showing sufficient evidence of validity and reliability**; the EPSRI may now provide novel insight on the moral reasoning of senior chemical engineering students. (2) **Contents Under Pressure, as a digitally immersive game, yielded decisions that were more authentic than those obtained from the EPSRI, as a traditional form of ethical assessment**. This outcome suggests that immersive digital games may provide a practice field that produces authentic opportunities of engagement similar to those of the real world which allow students to gain some practical experience free from real world hazards. (3) **Contents Under Pressure, as a form of pedagogy, found evidence of preparing senior chemical engineering students to make significant advancements to their moral development**. While traditional curriculum may lead students to predicting heightened ethical behavior, as an educational supplement, Contents Under Pressure encouraged morally transitional behaviors. This outcome means students reasoned with multiple levels of Kohlberg's framework in a step towards authentic post-conventional reasoning to consider concepts such as social justice or the environment. (4) **Senior chemical engineering students prioritize safety as a criteria in prediction, action, and reflection, and the personal relationships criteria became significantly more important through the process of gaining experience in decision making**. This outcome shows that safety is a priority throughout students' decision making. In contrast, plant productivity suffered in favor of safety. As such, process safety curricula need to incorporate conversations around competing criteria as not to discount their value. Also, students may have initially deprioritized personal relationships because engineering students lacked practical experience with making decisions around their relationships.

Contribution to Educational Research

This work developed the Engineering Process Safety Research Instrument (EPSRI) through reliability and validity testing, which provides engineering education researchers with a tool that is readily available to evaluate process safety ethics interventions. This work also provides researchers with a model of evaluating the effectiveness of a game-based education intervention, which specifically addresses shortcomings of other game-based learning studies as found by Bodnar et al. (2016).

Moreover, this work provides engineering education researchers with valuable insight on how senior chemical engineering students make process safety decisions in ethical dilemmas and in situations that have competing criteria. Ideally, this work contributes to the growing body of research on process safety decision making and human factors (Baybutt, 1996, 2017b, 2017a, 2015).

Contribution to Educational Practice

The EPSRI provides educators with an opportunity to evaluate the impacts of their own curriculum and with assistance in guiding their students to an awareness of the ethical implications of their decisions through reflection (Bairaktarova & Woodcock, 2017; Clancy, 2020a; Osberg & Shrauger, 1986). Specifically, the EPSRI may be used to assess changes in moral reasoning or identifying students' predominant form of moral reasoning. As a survey, the EPSRI may be used cross-sectionally around a process safety course or intervention or used longitudinally across the chemical engineering curriculum. The EPSRI may also be used in think-aloud studies, such as in other work by Bodnar et al. (2020).

In addition, this work evaluated a digital, process safety game that enabled students to practice their decision making in the context of ethical dilemmas and with complex competing criteria. Educators may easily incorporate this game to supplement their curriculum to build their students' decision making experience. For example, Contents Under Pressure may be incorporated as a multi-day participation-graded homework assignment or as a single-session play through as a class. Gameplay should be followed by in-class discussion around the challenges and ethical dilemmas faced in the game. Example discussion points may include:

- *How did the characters and the narrative of Contents Under Pressure influence your decision making?*
- *Through your decision making, were you able to obtain outcomes that aligned with your goals in the game?*
- *During your gameplay, what experiences surprised you to be easier or harder than you expected?*

- *After completing the game, what are some decision making strategies you would like to maintain or adopt?*

Applying both the EPSRI and Contents Under Pressure align well with ABET's desired student outcomes, which states that students must attain "*an ability to recognize ethical and professional responsibilities in engineering situations and make informed judgments, which must consider the impact of engineering solutions in global, economic, environmental, and societal contexts*" (ABET, 2021b).

Summary

The chemical process industry has increasingly invested in process safety education and preventative measures, yet many efforts have overlooked the role of an engineer's decision making. As a result, decision making choices have contributed to process incidents, and additional incidents will likely continue to occur without further intervention. As such, this work promotes a better understanding of process safety decision making amongst senior chemical engineering students through three studies.

Specifically, this work completes development of the EPSRI to assess students' moral reasoning in process safety decisions. Work completed has shown that the EPSRI presents sufficient evidence of validity and reliability for use with senior chemical engineering students. This work also evaluates the impact of a digital, process safety game on engineering students' moral reasoning by comparing rated importance of potential considerations and by comparing EPSRI scores between intervention and comparison cohorts. These comparisons found that responses made within the game setting better aligned with anticipated moral reasoning results (Rest, Thoma, & Edwards, 1997) than responses to the considerations within the EPSRI. Moreover, students who played the game during the semester may have been more prepared to make advancements to their moral reasoning. Last, this work explored how students prioritized criteria through their decision making when in situations with competing criteria. Students provided evidence of primarily prioritizing safety above all other criteria in predictive and reflective surveys and in practice within CUP. Yet, this approach in-game frequently led to game “failures” from the plant output metric operating at unacceptable levels.

The following sections provide a review of relevant literature (Chapter 2), development of the EPSRI (Chapter 3), evaluation of the Contents Under Pressure game (Chapter 4), and investigation of competing criteria on decision making (Chapter 5).

Chapter 2

Review of Literature

Introduction

This chapter is structured around four key sections. First, it extends the literature from Chapter 1 to provide additional detail on problems within process safety decision making and potential pedagogical solutions. This section will also review blind spots to ethical decision making, particularly behavioral forecasting and ethical fading. Subsequently, the theoretical framework used in understanding the ethical principles behind decisions considering three levels of moral development will be introduced in the second section. The third section will explore the small but growing body of literature on how competing criteria may influence decision making where ethical implications may be non-obvious. Finally, the last section of the chapter will briefly review the benefits of digital games and their potential implications towards education as they can be useful practice fields for providing firsthand experiences in an environment that is safe from actual process safety hazards.

Process Safety Decision Making and Blind Spots

Decision making is a critical skill for engineering practitioners. The Accreditation Board for Engineering and Technology (ABET), who defines the standards and criteria that make up undergraduate engineering programs, requires that engineering programs produce engineers who can use their judgement to make balanced and ethical decisions in preparation for professional practice (ABET, 2021b). In addition, decision making is critical because of its capacity to inhibit engineering design (CSB, 2015, 2017). In chemical process safety, decision making often creates opportunities to bypass layers of protection. Ness (2015) visualizes this process through the Swiss cheese model (Figure 2.1). The left side of the visual shows process operations that are protected from divulging into incidents through layers of protection (Ness, 2015). These layers may include hazard analyses (HAZOP), layers of protection analysis (LOPA) (Cong et al., 2016; Willey et al., 2020), warning alarms, personal protective equipment, and written protocols (Crowl & Louvar, 2019). However, these layers of protection are like slices of Swiss cheese: full of holes that limit their viability. Where these holes align, an incident may occur (Ness, 2015). Process engineers acknowledge these pathways to failure, and thus, aim to add more layers of protection (Crowl & Louvar, 2019). However, in some cases, decision making can create an incident pathway by making new holes within the layers of protection.

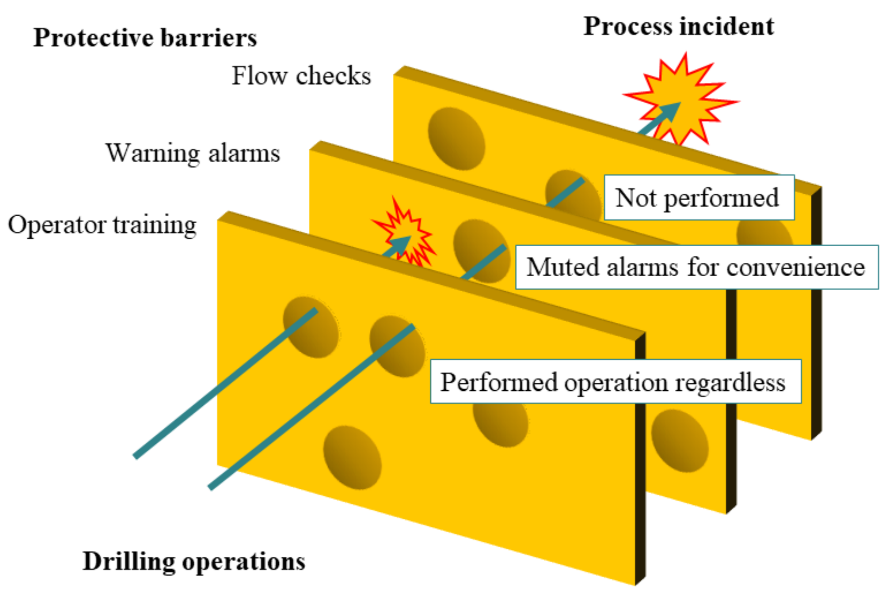


Figure 2.1. The Swiss cheese model illustrates how human decision making can inhibit layers of protection, such as from the Pryor Trust case study. Adapted from Ness (2015).

For example, in 2018, a natural gas well, Pryor Trust 1H-9, experienced a catastrophic blowout and drilling rig fire that fatally injured five workers (CSB, 2019b; Money, 2020). A well blowout occurs when drilling pressure control systems fail, releasing flammable natural gas into the surrounding environment (CSB, 2019b). During drilling operations, as with other processes, engineers protect against incidents through multiple barriers (Ness, 2015). Drilling operations at the Pryor Trust well included barriers such as training workers on the equipment specific to such an operation, following pre-established safety procedures (i.e., flow and pressure checks), and acknowledging warning signs designed into the system. Human decision making may have contributed to why these barriers failed. Instead of training on the specific equipment, operators drilled regardless of lacking proper training. Workers did not perform flow and pressure checks, leading to an over pressurized well. Last, multiple workers muted alarms warning of the imminent blowout (CSB, 2019b). The end result was operators at the Pryor Trust well making multiple decisions, as illustrated in Figure 2.1, that broke through protective barriers, which led to a fatal well blowout (CSB, 2019b).

When avoiding potential process failures, it becomes increasingly important to understand how decisions are made because these decisions play such an influential role in engineering process operations. While the Pryor Trust case study has some decisions that may have more evident red flags of poor decision making (e.g., muting alarms) than other process safety case studies, the ethical implications of decisions are not always clear in the moment. Lack of clarity may come from blind spots to ethical implications.

The following literature review describes two factors that create blind spots in engineering decision making.

Behavioral Forecasting

In hindsight, poor decisions, such as muting the safety alarms (CSB, 2019b), may appear obvious. Yet, believing one would never make such decisions is a form of behavioral forecasting. Behavioral forecasting is where one predicts how they will behave, react, think, or feel in a specific situation while removed from that situation (Bazerman & Tenbrunsel, 2011). While behavioral forecasting is not inherently unethical for decision makers, inaccurate forecasts may create ethical "blind spots" that overshadow actual behaviors (Bazerman & Tenbrunsel, 2011). Osberg and Shrauger (1986) suggest that in these blind spots unethical behavior is most likely to arise. Specifically, individuals may consciously justify predictions to an ethical framework; however, when their actual behaviors deviate from these predictions, their behaviors are less likely to be justified by ethics. It is these behaviors, which are caused by inaccurate predictions, that are of concern because of their potential to be unethical. Business ethics researchers Bazerman and Tenbrunsel (2011) cite an example study of this in their work. In this study, women were asked to predict how they would respond to sexually harassing questions during a job interview, and many wrongly predicted how they would respond once put into that position in a lab setting (Woodzicka & LaFrance, 2001). The difference between these individual's predictions and practice highlights a blind spot within their ethical behaviors. Inaccurate predictions, such as the individuals in Woodzicka and

LaFrance’s (2001) study, may stem from inexperience in decision making or ethical fading.

Ethical Fading

Ethical fading is a psychological phenomenon when the ethical elements of a decision are blurred among the multitude of factors and responsibilities that contribute to a decision. Tenbrunsel and Messick (2004, p. 224) first introduced the terminology ethical fading as the “process by which the moral colors of an ethical decision fade into bleached hues that are void of moral implications.” This fading may lead to individuals misclassifying a decision to be a typical or mundane decision opposed to a critical ethical decision. Inability to differentiate the two may lead to a blind spot where one needs to apply ethics to their decision making (Bazerman & Tenbrunsel, 2011).

Testino (2007) posits that ethical fading is a function of an individual’s intuition. Thus, when not deliberately examining every situation for ethical implications, they are often overlooked. Bazerman and Tenbrunsel (2011) exemplify Testino’s point by suggesting that the Space Shuttle Challenger disaster (McDonald & Hansen, 2009; Rogers et al., 1986) occurred because of ethical fading. In this event, Morton Thiokol’s engineering decision makers were alerted that an engine O-ring may fail during the launch. Rather than examining the ethical elements of the situation, the engineers claimed they needed to make a “management decision” (McDonald & Hansen, 2009). Moreover, Clancy (2020a) acknowledges that ethical implications are easily perceived when they are expected to be there. Clancy et al. (2017) suggests that to mitigate fading engineering students should become familiar with ethical implications and ethical framing. Thus, gaining experience can mitigate unethical decisions caused by this blind spot.

Chemical engineering educators and practitioners agree that decision making is crucial to the profession (ABET, 2021b; Baybutt, 1996, 2017b, 2017a, 2015; Crowl & Louvar, 2019). Despite community agreement, engineers are not immune to ethical blind spots (behavioral forecasting and ethical fading) in their decision making. In response to the risk of these blind spots, multiple researchers advocate for building awareness of one's own ethical behaviors and of the ethical implications of one's decisions (Bairaktarova & Woodcock, 2017; Clancy, 2020a; Osberg & Shrauger, 1986). This approach may be particularly helpful in process safety dilemmas where process engineers' decisions are far removed from their ethical implications, such as impacts to surrounding communities or to the environment. As engineering educators develop forms of decision-making education and curricula to address ethical awareness, educators must also develop forms of assessment.

Kohlberg's Theory of Moral Development

Foundations

Kohlberg's theory of moral development offers a potential foundation for assessing decision making education. Kohlberg's theory of moral development as described by Kohlberg & Hersh (1977) consists of three levels of reasoning: pre-conventional, conventional, and post-conventional. Individuals who reason pre-conventionally may be motivated by personal interests, such as rewards or selfpreservation. Pre-conventional reasoning is associated with maintaining one's public image, promoting one's personal finances, or conserving one's time. Individuals who reason conventionally may be motivated by those in their immediate surroundings, instead of just themselves. Conventional reasoning expresses concern for employer objectives and for the people known personally, such as family, friends, or coworkers. Last, individuals who reason post-conventionally may be motivated by philosophical ideas, such as justice and goodwill. Actions based in this level of reasoning may benefit society as a whole or the environment (Kohlberg & Hersh, 1977).

Moral reasoning is a developmental model, meaning humans start at the lowest level, pre-conventional, and advance to higher levels, conventional, and post-conventional. The model is non-regressive, meaning when higher levels of moral reasoning are obtained, the preceding lower levels are still accessible to individuals because (Kohlberg, 1985; Kohlberg & Hersh, 1977). Kohlberg first understood these levels to be distinct and similar to stairs with discrete steps in development (Kohlberg, 1981; Kohlberg & Hersh, 1977). Yet, neo-Kohlbergian advancements to the theory acknowledge that moral development gradually shifts slowly between the levels (Rest et al., 2000; Rest, Narvaez, Bebeau, et al., 1999; Siegler, 1997). Using this advancement, moral reasoning is better described through six *schemas*, which are either consolidated or transitional to a Kohlbergian level (Rest, Narvaez, Bebeau, et al., 1999). Moral development, using both Kohlbergian and neo-Kohlbergian approaches, is illustrated in Figure 2.2.

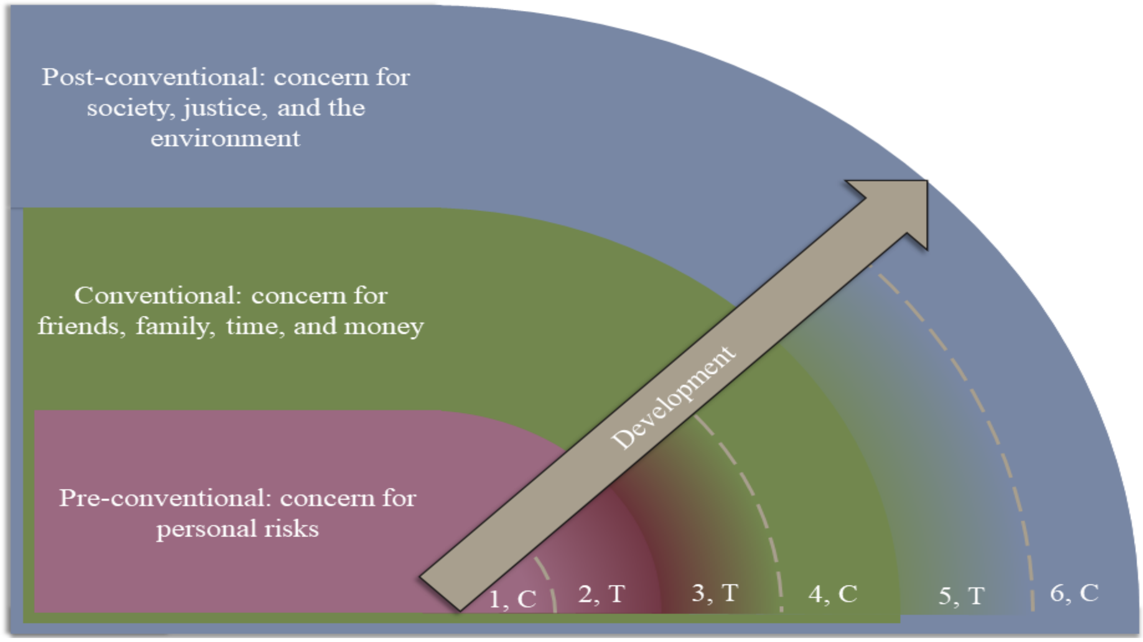


Figure 2.2. Visualization of moral development with both discrete Kohlbergian levels and gradient neo-Kohlbergian schemas. The six schemas may either be consolidated (C) or transitional (T) to a level of moral reasoning.

Neo-Kohlbergian schemas describe how consistently one reasons at a Kohlbergian level. Thus, consolidation describes how an individual predominantly reasons at a given level, and transition describes how an individual reasons primarily using a blend of two levels (Rest, Narvaez, Bebeau, et al., 1999). Transitional moral reasoning exists where one level is preferred, yet reasoning with other levels may exist.

Appropriateness of Moral Reasoning

The neo-Kohlbergian approach to moral reasoning is a suitable model for this work because it allows observation of the principles behind decisions. The level of moral reasoning an engineer uses speaks nothing of the correctness of the decision because many times there are not ethically "right and wrong" answers, and the most appropriate response may be contextual to the dilemma. As an example, the trolley dilemma is a common ethics problem where the right answer is debated by the ethical perspective taken (Shafer-Landau, 2018).

In a version of the trolley dilemma, a runaway trolley races towards a crowd who will certainly be killed on impact. The reader of the dilemma has the opportunity to throw a lever that will send the trolley down a parallel track that will kill only a single individual. Deciding to pull the lever might employ utilitarian ethics with the goal of doing the most good or saving the most lives (Mandal et al., 2016). In contrast, deciding not to pull the lever might employ deontological ethics with the goal of being as fair and just as possible (Mandal et al., 2016). Pulling the lever would be sentencing the single individual to their death, which would be unfair to them. Ongoing philosophical debates suggest that neither response is "right" nor "wrong." Thus, moral reasoning evolves as a suitable theoretical framework because it allows us to observe what principles are applied to decisions without engaging in the argument of ethical correctness. Because this model is developmental, some engineering education researchers aim to advance engineering students to post-conventional schema to focus on ethical justice (Clancy, 2021). To differentiate from neo-Kohlbergian moral reasoning, which focuses on principles of ethics, these researchers use the term ethical reasoning instead of moral reasoning. While it may be desirable to advance to higher levels of moral reasoning, that is not the immediate purpose of these studies.

Applications of Moral Reasoning

Despite moral and ethical reasoning frameworks having discrepancy in the approach to development, these frameworks have widely been accepted within engineering education as guiding theories or embedded within instruments. Early efforts to measure moral reasoning lead to the development of the Defining Issues Test (DIT) and the Defining Issues Test, Version 2 (DIT2) (Rest, 1974; Rest, Narvaez, Bebeau, et al., 1999; Rest, Narvaez, Thoma, et al., 1999; Rest, Thoma, Narvaez, et al., 1997). These instruments found evidence of validity through a large sample from the general population. These tests' formats present respondents with multiple ethical dilemmas, each with two viable solutions and a third choice of "can't decide." Next, the tests ask respondents to rate a series of potential *considerations* to their decision and rank their top four (Rest, 1974; Rest, Narvaez, Thoma, et al., 1999). Instrument developers design each of these considerations to represent the three levels of moral reasoning. For the DIT and DIT2, an example dilemma may be whether the respondent would steal life-saving medicine from a pharmacy if they could not afford it. A conventional designed consideration may ask respondents if they considered the impact of their decision on the lives of the pharmacy owner, and a post-conventionally designed consideration may ask if they considered implications of crime on the community (Rest, 1974; Rest, Narvaez, Thoma, et al., 1999). The rating and ranking of these considerations result in a quantitative measure of the respondent's moral reasoning.

Borenstein et al. (2010) out of the Georgia Institute of Technology, attempted to use the DIT2 in their study on science and engineering students. However, they were unable to find practical changes from an ethics intervention possibly because the DIT2 was not sensitive to nuances of engineering and science disciplines. As such, Borenstein et al. developed the Engineering and Science Issues Test (ESIT), a new moral reasoning instrument tailored towards their sample. Since its development, the ESIT has been used as a tool to evaluate ethics courses (Selby, 2015), understand cultural influences on ethical reasoning (Clancy, 2020), and personalize ethics lectures based on students' ethical reasoning (Kerr, 2016). Similar to the development of the ESIT, a study by Titus et al. (2011), out of Purdue University, observed that engineering students did not find the DIT2 dilemmas to be relevant to their engineering education. In response, these researchers developed the Engineering Ethics Research Instrument (EERI) (Zhu et al., 2014). Despite the EERI lacking evidence of validity, it has been used to evaluate the effectiveness of ethics interventions (Ghorbani et al., 2018; Murzi et al., 2019; Reed, 2022; Reed et al., 2021). Both the ESIT and the EERI were developed to capture nuances in engineering ethical decisions. Thus, the development of and practice with these two instruments in engineering contexts promotes the appropriateness of assessing engineering ethical decision making through a Kohlbergian-based moral reasoning instrument.

Framework Limitations

Osborne (2012a) acknowledges that researchers can never obtain a "perfect" sample. Instead, researchers should use theory to drive sampling decisions and should exhibit transparency on limitations. Kohlberg's theory of moral development may be limited in its generalizability among gender because it was developed with a predominantly male sample (Rest, Thoma, & Edwards, 1997). Gilligan and Attanucci's (1988) work highlight studies suggesting there may be differences in moral reasoning by gender, albeit no meaningful differences can be claimed because these studies were inadequately powered and did not report effect size. Moreover, there is no theoretical reason suggesting that there should be differences in moral reasoning by gender, but this limitation must be acknowledged.

Competing Criteria in Decision Making

In addition to moral reasoning, decision making can also be understood through the competing criteria practitioners may encounter. Criteria refers to metrics or principles, which may compete for attention or valuation in the eyes of a decision maker. These criteria may include 1) budget, 2) personal relationships, 3) plant productivity, 4) time, and 5) safety. Criteria are a meaningful influence to decision making based on literature (Akinleye et al., 2019; CSB, 2014; Dönmez & Uslu, 2020; Hendrickson & Au, 2008; U.S. Chemical Safety and Hazard Investigation Board, 2018) and informal feedback from students who played a digital process safety game. Competing criteria offer a lens that must be looked through to holistically understand engineering decision making because not all decisions have obvious ethical elements. As such, engineers may evaluate decisions and the outcomes by the criteria tied to the responsibilities of the profession and to the interests of its practitioners. Some of these criteria have been considered in the past through decision making literature (Howard & Abbas, 2016). Yet, these past approaches focus on quantitatively maximizing or minimizing scale variables. However, this approach overlooks time-sensitive or intuition requiring decisions that process engineers must face. Because the competing criteria lens is novel, literature is limited in describing how engineers approach decision making from the need to balance criteria, especially among chemical process engineers. As such, this work employs literature on competing criteria from process safety case studies and other industries to develop a conceptual framework around five competing criteria (budget, personal relationships, plant productivity, time, and safety).

When making decisions, process engineers may need to consider budget as a criteria. This criteria has to do with adhering to company budgets and minimizing spending to yield a higher net profit. In practice, engineers at DuPont designed a phosgene transfer system with cheap, short-life hoses because of the potential cost savings over time (CSB, 2011), yet such decisions may impact worker safety in favor of budget. This design decisions forced workers to replace the hose more frequently, putting them at a higher risk of chemical exposure, which ultimately led to a fatal exposure after a hose failure (CSB, 2011). Crowl and Louvar (2019) acknowledge this balance between budget and safety, and they promote investing in process safety. They challenge the myth that process safety may negatively affect a company's economic performance by reflecting on the success of Alcoa who are leaders in both aluminum production and process safety (Crowl & Louvar, 2019). An example of failing to manage competing criteria recently played out at the Philadelphia Energy Solutions refinery, where a catastrophic fire erupted from a pipe leak that had been flagged for replacement over a decade ago (CSB, 2019a). After the pipe had been flagged for thinning, it is likely this segment of pipe had not been replaced due to budget constraints, affirming the difficulty of balancing budget and safety criteria.

Process engineers may also need to consider personal relationships in their decision making. Personal relationships take into account perception and opinions of coworkers, and connections with them or others who may be important, such as family and friends. These relationships can affect engineers' perception of safety culture and professional responsibilities. For example, faltering safety culture at an Aghorn oil pump station in Odessa, Texas promoted a pump worker's spouse to fatally enter a restricted work site against safety protocols during a gas leak (CSB, 2021). Another incident report from the CSB (2015) identified that upon engineers identifying a hazardous chemical leak and agreeing the plant needed to be shut down to minimize risks, the plant remained in operation until a fireball erupted from the leak. While the engineers had deliberated and agreed on the need to halt operations, all of the engineers involved neglected to make the call, some of whom reported that they did so to avoid "flack" for making the call (CSB, 2015, p. 81). In both of these cases, engineers may have been victim of affect bias, which is where emotions and feeling, such as fear or pleasure, influence a decision (Baybutt, 2017b). Paul Baybutt, an industry expert in process safety and risk assessment, emphasizes that everyone, including engineers, can fall victim of affect bias (Baybutt, 2017b, 2017a). Therefore, engineers must consider personal relationships (and the implications of prioritizing personal relationships) through their decision making.

Plant productivity refers to the bottom line and production demands of customers and employers. Process engineers consider plant productivity in process safety decisions, as it is engrained in their decision making from undergraduate education. For example, ABET requires chemical engineering students to be knowledgeable in analyzing hazards (ABET, 2021b; Vaughen, 2012). In response, many chemical engineering programs added hazard awareness to their curriculum, such as through hazard and operability studies (HAZOP) (Willey et al., 2020). Crowl and Louvar (2019) note that HAZOP focuses on understanding process hazards and their *impacts on productivity*. Thus, the foundations of process safety and analyzing hazards are tied to plant productivity as a criteria.

Time as a criteria has to do with an engineer's availability, whether that be for personal interests or career advancement. Time may also be considered in process safety decisions and is evident in process incident case studies. In 2016, unit operators at an ExxonMobil refinery were tasked with closing a valve, but the operators found the valve gearbox was jammed. Breaking protocol in preference of saving time, the operators made the decision to disassemble the valve gearbox, which resulted in a fatal isobutane leak (CSB, 2017). In another case study, engineering management scheduled welding repairs of a pipe segment near a condensate tank. However, leadership dedicated insufficient time to evacuate flammable fumes from the tank, which resulted in an explosion that killed two maintenance welders (CSB, 2018). In both of these situations, time was not appropriately dedicated to the task at hand, which lead to undesirable outcomes.

Last, process engineers must consider safety as a criteria to their decisions. Safety encompasses wearing personal-protective-equipment, preventing employee injury, protecting plant machinery, and safe-guarding the environment and surrounding communities from chemical leaks (Crowl & Louvar, 2019). In an incident investigation, the CSB (2019c, p. 81) quotes DuPont's Process Safety Management policy, "-where safety priorities are recognized, not as conflicting with other priorities, but rather as inherently necessary for completing any task the right way." Furthermore, ABET requires that engineers be able to consider public health, safety, and welfare in every engineering solution (ABET, 2021b). While safety may appear to compete for attention with other criteria in prior examples, it is not meant to conflict with others. Safety must be a consideration to any process safety decision.

While safety is inarguably important to the process industry, safety is linked to other criteria and often competes with them for priority. Examples of this behavior are found when safety is neglected by affect bias (Baybutt, 2017b, 2017a) in favor of production demands at the Chevron refinery (CSB, 2015), when safety is neglected for convenience or to save time (CSB, 2018), or when safety is tied to production and budget concerns (Crowl & Louvar, 2019; CSB, 2011). Competing criteria are not unique to the process industry. Studies among hospital nurses found patient care and well-being (<u>safety</u>) deteriorates when competing with hospital <u>budget</u> constraints (Akinleye et al., 2019; Encinosa & Bernard, 2005). Again, in the airline industry, incidents (<u>safety</u>) occur in-part due to poor <u>leadership</u> from air traffic controllers (Dönmez & Uslu, 2020). Hendrickson and Au (2008) acknowledge dedicating <u>time</u> to planning at the forefront of a construction or civil engineering project is critical to reducing expenses (budget) and on the job injuries (safety). Practitioners, engineers included, need to balance these criteria, and be aware of impacts that may occur when over prioritizing or neglecting a criteria. Despite evidence that the competition of criteria may influence engineers' decision making, there is limited literature describing *how* engineers actually consider or incorporate criteria into their decision making.

Game-Based Learning

Foundations of Game-Based Learning

Games have been treated as an umbrella term to describe a multitude of systems and contexts including tools, media, experiences, art (Schrier, 2021), virtual realities, and simulations (Pasin & Giroux, 2011). Ma et al. (2011; 2017) uses the term "serious game" to describe digital-educational games. In addition, while acknowledging the lack of a formal definition, Dörner et al. (2016) describes serious games as digital games designed to achieve both entertainment and one other goal, such as learning or exercise (Lumosity, Wii Fit, etc.). Kapp (2012) uses attributes of games to broadly describe them; these attributes include challenges with rules, systems of player inputs and game feedback, abstraction, interactivity, quantifiable outcomes, and emotional reactions. Crawford (1997), an expert of computer game design, explains games must also include conflict, yet despite the conflict, players must be safe from the dangers of conflict. Dörner et al. (2016) expands this list of attributes to include storytelling, social factors, and learning. Games are unique from simulations because games may also engage players through narrative, reward systems, and non-playable characters (NPCs) (Young et al., 2012). By incorporating these viewpoints, this work interprets games to be systems of play and enjoyment, bound by rules and challenging objectives, with controllable actions and feedback unique to those actions, and with outcomes that are measurable and that influence the player's emotions. Moreover, a serious game is a digital product, through which learning may occur.

Despite the vast approaches to games, the simplest understanding of games is that they are "play" structured by rules (Crawford, 2003). Many have taken this structure and purposefully designed games as a form of pedagogy known as game-based learning (GBL) (Ma et al., 2011). GBL is becoming more prevalent as literature reviews are being performed to evaluate their effectiveness in the classroom (Bodnar et al., 2016; Young et al., 2012). Educational games have been applied to all stages of education in the US: K-12 (Vogel et al., 2016), higher education (Harris et al., 2009; McDaniel et al., 2007), and industry training (Juego-Studio, 2023; Microsoft, 2013). In addition, digital game pedagogy has also become more prevalent within classrooms and multiple industries as technology has become more commonplace (Ma & Oikonomou, 2017). While Young et al. (2012) acknowledges there is no practical repository of educational games that have been evaluated for their effectiveness, the benefits of games on pedagogy have been well documented.

A major benefit of games toward play or education is immersion, where the mediators and barriers between a player and the game fade out of existence (Przybylski et al., 2010). Immersion shares reactions similar to Csikszentmihalyi's theory on the state of flow. Flow occurs when an individual is concentrating so deeply on an activity that they lose focus of time and their surroundings; the individual often finds the activity they are engaged in to be intrinsically enjoyable and performs the activity most optimally (Nakamura & Csikszentmihalyi, 2014). Immersion in digital games evolves from three domains grounded in the self-determination theory of motivation (Ryan et al., 2006); these domains are physical presence, emotional presence, and narrative presence (Przybylski et al., 2010).

Two primary factors drive immersion: gaming interface and player's needs. Users interface with games through controls, graphics, sounds, and haptic feedback. Johnsen et al.'s (2018) pilot study on an educational nursing game found that the game's controls were too simplistic giving their players not enough control of their actions. On the other extreme, Orvis et al. (2010) found that too complex of controls can overwhelm players encouraging them to disengage from gameplay. Either of these extremes promote barriers of immersion (Nakamura & Csikszentmihalyi, 2014), so optimized gameplay requires balanced control complexity. When discussing educational game design, Mildner and Mueller (2016) acknowledge how players are able to control a game, see the game, and even save their progress may have an influence on their immersion. In contrast to designing immersive interfaces, Przybylski et al. (2010) found that meeting the player's motivational needs through narrative is among the strongest drivers of immersion. This claim is consistent with the findings of Bormann and Greitemeyer (2015) who found that narrative in digital games drove immersion that could positively influence games' learning outcomes.

Educators acknowledge that classroom experiences may be inauthentic to real world experiences leaving students ill-prepared, but immersion in digital games may address these concerns (Nicaise et al., 2000). While education researchers have called for more hands-on training, this may be increasingly difficult in high risk professions, such as medicine or process engineering. Educational games provide an opportunity to fill this need. First, immersion through games can create authentic environments (Mildner & Mueller, 2016; Newstetter & Svinicki, 2014) where players may act in authentic ways; as such, immersion in games can remove predictive biases and promote ethical fading (Bazerman & Tenbrunsel, 2011) unlike in-person classroom experiences. In addition, games permit players to simulate “reality” without real world repercussions (Crawford, 1997; Hejdenberg, 2005). This benefit has been leveraged in serious game trainings, such as flight simulators (Dörner et al., 2001), nursing practice (Johnsen et al., 2018), or public safety (Martínez-durá et al., 2011). As such, educational games may be used as “practice fields” where players can build up practical experience without legitimate dangers (Barab & Duffy, 2012; Crawford, 1997).

While there are significant benefits of immersion, serious games may have limitations on who they can reach. General social perception suggests that digital games are a form of entertainment for men instead of women and acknowledges that women may be less interested in playing games than males (Orvis et al., 2010). However, Joiner et al. (2011) directly investigated mechanical engineering students for gender differences in knowledge and motivation around an educational, digital game and found no significant differences. Another concern is the role of player age. Orvis et al. (2010) explored the value in educating military personnel by analyzing over 10,000 responses to the Sample Survey of Military Personnel, and they found significant decreases in gameplay experience with age. Because this finding identifies a potential risk to educational game effectiveness, they call for additional longitudinal data to differentiate generational differences from age differences.

Despite questions on the role of demographics in game-based learning, both Pasin and Giroux (2011) and Johnsen et al. (2018) support that games supplement traditional course curriculum as it provides multiple opportunities to learn. Pasin and Giroux (2011) described this approach to be effective because it allowed students to help grasp complex topics with firsthand experience. Tawa (2017) echoes this effectiveness with an example of a serious game that was used in a cross-cultural psychology course on racial differences, which was noted to be achievable only through the immersive nature of the game. The aforementioned studies by Johnsen et al. (2018) and Tawa (2017) may have been found to be so successful because of their ability to immerse their players into the game.

Examples of Game-Based Learning

In relation to this work, literature outlines three relevant GBL domains as they relate to ethical process safety decision making: safety, awareness building, and discipline-specific type games. Some games are used to teach safety related to high risk environments where it may be unrealistic to facilitate training or see the outcomes of poor practice. Games may also be used to build awareness in social or ethical dilemmas. These games provide players with an opportunity to view dilemmas from new perspectives or gain understanding through simplicity of play. In addition, certain disciplines have begun using games because of their ability to teach complex ideas and skills. A summary of examples in these domains and their overlaps are shown in Figure 2.3. The overlap of these three GBL domains defines the attributes needed in a game that can be used to teach ethical process safety decision making. Thus, the game selected for this study is Contents Under Pressure, which will be described in detail in Chapter 4 where it is first applied within a study.

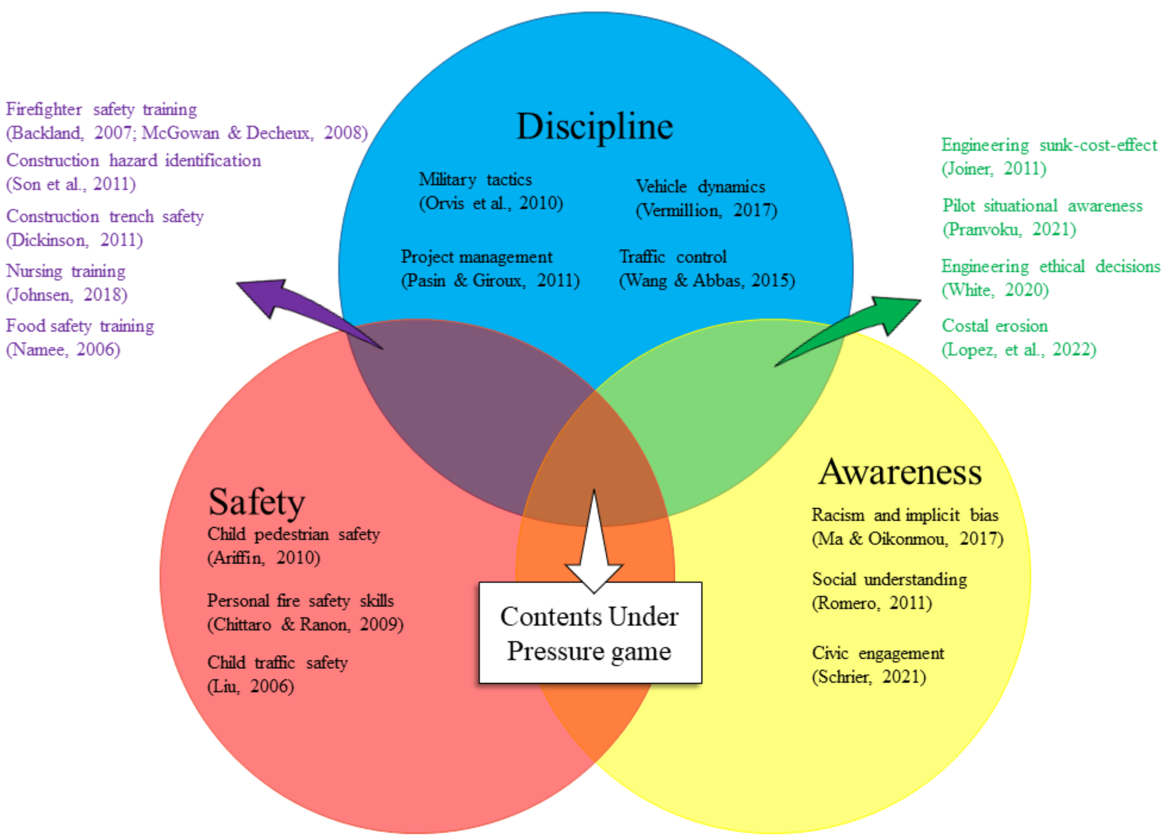


Figure 2.3. Summary of literature among three GBL domains where the overlap contextualizes Contents Under Pressure.

Summary

This literature review covered concepts related to decision making, moral reasoning, competing criteria, and game based learning. Engineering decision making plays a critical role in projects and social well-being, yet this role may be under investigated in comparison to other approaches to safety and design. This may create a deficit in decision making experience, leaving openings for bias and confounds, such as behavioral forecasting and ethical fading, to influence engineering decisions. Chemical engineers are not uniquely immune to these biases. Kohlberg's theory of moral development may be used to evaluate the principles behind these decisions without determining an ethically right or wrong response to a dilemma. Moreover, ethical elements to everyday decisions are not always obvious, and criteria may influence engineers' decision making. This literature review identified five potential competing criteria that engineers may consider: budget, personal relationships, plant productivity, safety, and time.

Last, game-based learning may leverage attributes of immersion to provide students an opportunity to train in an authentic manner without penalties and repercussions of the real world. This is applicable to process safety education because classrooms may not provide authentic experiences where engineers can build an awareness of their own decision making regarding ethics and competing criteria.

Chapter 3

Developing the Engineering Process Safety Research Instrument

Overview

The objective of this first study was to complete the development validation of the Engineering Process Safety Research Instrument (EPSRI) through exploratory (EFA) and confirmatory factor analyses (CFA). Instrument development validation was begun in earlier work (Butler, 2018) and was completed in this study. Multiple process incident case studies supply evidence that they were caused by a series of poor engineering decisions (CSB, 2014, 2019b). Literature suggests that strengthening ethical decision making education holds the potential to mitigate such incidents by building awareness and experience in such decisions (Bairaktarova & Woodcock, 2017; Clancy, 2020a; Osberg & Shrauger, 1986). However, up to this point, no form of ethical decision making assessment exists in process safety contexts to test such educational endeavors, which may hinder ethical decision making education in this context. As such, this work sought to establish the EPSRI as an assessment tool of moral reasoning for senior chemical engineering students focused on the context of process safety.

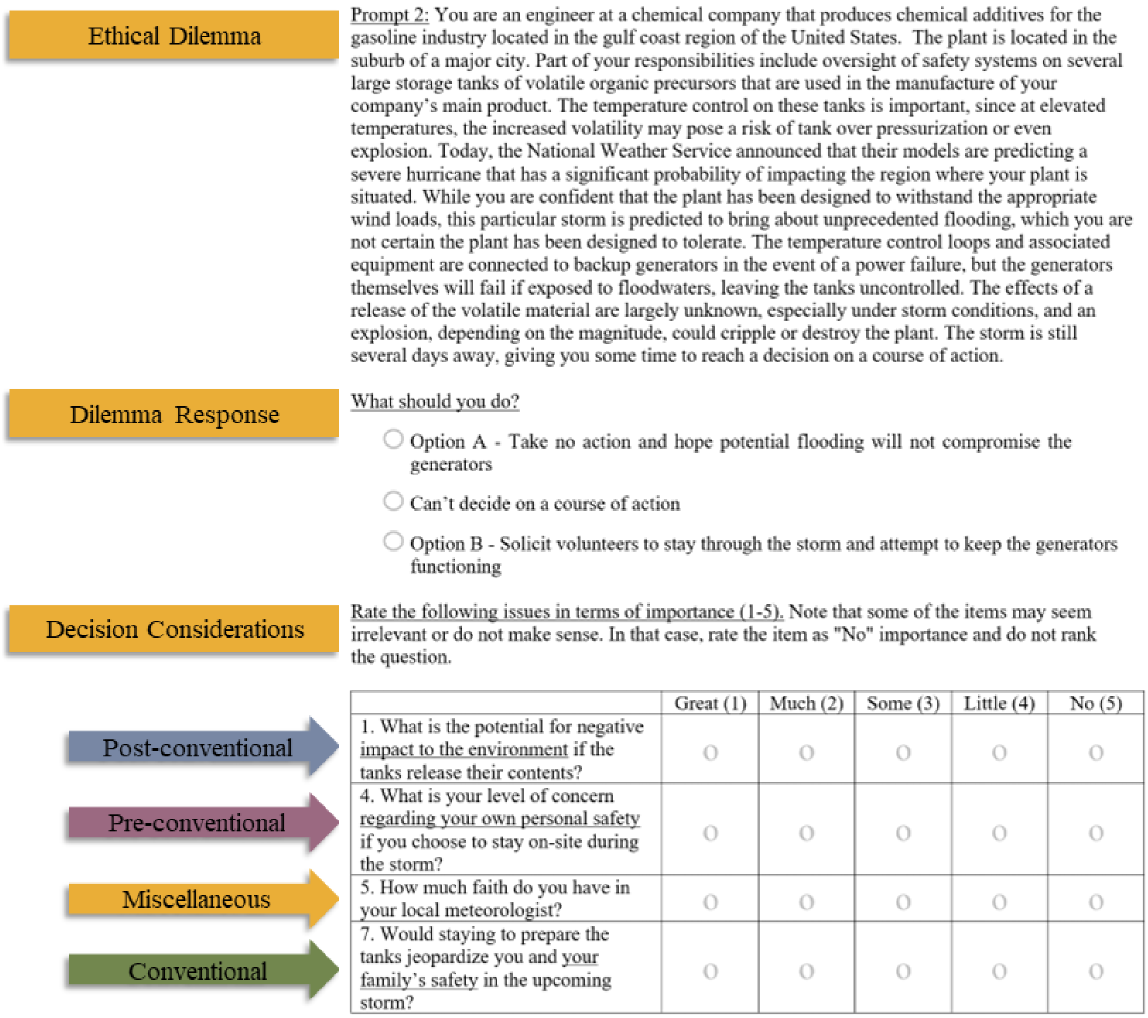
Modeled after the seminal moral reasoning instrument, the DIT2 (Rest, Narvaez, Thoma, et al., 1999), the EPSRI presents respondents with a series of ethical dilemmas, possible responses, and a list of items that they may have considered while making each decision (considerations). Through an EFA, this study sought to find evidence that these considerations represent the intended levels of moral reasoning (pre-conventional, conventional, and post-conventional (Kohlberg & Hersh, 1977)) through statistical correlations (validity) and that these correlations exist consistently in the data (reliability). Subsequently, this study performs a CFA to affirm the developed instrument and correlations among the considerations are a “good fit” for the data (Floyd & Widaman, 1995, p. 293). This study did find some evidence of validity and reliability of the EPSRI as a tool to assess the moral reasoning among senior chemical engineering students.

Introduction

Moral Reasoning Assessment

Moral reasoning is an intangible trait, which has historically been measured through the Defining Issues Test (DIT) and the Defining Issues Test 2 (DIT2). However, engineering education researchers recognized that engineering students did not find relevance to engineering from the general dilemmas posed in these instruments. In response, researchers from Georgia Institute of Technology and Purdue University responded by developing the Engineering and Science Issues Test (ESIT) (Borenstein et al., 2010) and the Engineering Ethical Reasoning Instrument (EERI) (Zhu et al., 2014). These examples emphasize that moral development is contextual to each industry. Furthermore, the moral development of chemical engineers concerned with process safety is unique to each students’ own engineering experiences. For example, the ethical dilemmas that a civil engineer faces may differ from those that a biomedical engineer does. This progression led to the inception of the EPSRI as an assessment instrument capable of capturing the nuances associated with process safety’s ethical dilemmas. Prior development of the EPSRI is documented elsewhere (Butler, 2018).

The EPSRI is formatted similarly to the DIT2. Instrument respondents read a number of independent ethical dilemmas, choose from three potential solutions to each (one being “can’t decide”), rate the importance of a series of considerations relevant to each decision, and rank their top-four considerations. Dilemmas in the EPSRI are contextual to process safety, pulling from process incident case studies (Butler, 2018). The considerations were designed to reflect the three levels of moral reasoning from Kohlberg’s moral development theory (Kohlberg & Hersh, 1977). On occasion, meaningless considerations (M-items) would be included among the considerations that would clearly not apply to the dilemma. For example, an M-item might ask the importance of being able to order pizza while deciding how to operate a plant during a hurricane. These M-items were used to filter out bogus responses as described elsewhere (Butler et al., 2019; Rest, Narvaez, Thoma, et al., 1999). An example dilemma from the EPSRI is shown in Figure 3.1.



Ethical Dilemma

Prompt 2: You are an engineer at a chemical company that produces chemical additives for the gasoline industry located in the gulf coast region of the United States. The plant is located in the suburb of a major city. Part of your responsibilities include oversight of safety systems on several large storage tanks of volatile organic precursors that are used in the manufacture of your company's main product. The temperature control on these tanks is important, since at elevated temperatures, the increased volatility may pose a risk of tank over pressurization or even explosion. Today, the National Weather Service announced that their models are predicting a severe hurricane that has a significant probability of impacting the region where your plant is situated. While you are confident that the plant has been designed to withstand the appropriate wind loads, this particular storm is predicted to bring about unprecedented flooding, which you are not certain the plant has been designed to tolerate. The temperature control loops and associated equipment are connected to backup generators in the event of a power failure, but the generators themselves will fail if exposed to floodwaters, leaving the tanks uncontrolled. The effects of a release of the volatile material are largely unknown, especially under storm conditions, and an explosion, depending on the magnitude, could cripple or destroy the plant. The storm is still several days away, giving you some time to reach a decision on a course of action.

Dilemma Response

What should you do?

- ○ Option A - Take no action and hope potential flooding will not compromise the generators
- ○ Can't decide on a course of action
- ○ Option B - Solicit volunteers to stay through the storm and attempt to keep the generators functioning

Decision Considerations

Rate the following issues in terms of importance (1-5). Note that some of the items may seem irrelevant or do not make sense. In that case, rate the item as "No" importance and do not rank the question.

		Great (1)	Much (2)	Some (3)	Little (4)	No (5)
Post-conventional	1. What is the potential for negative impact to the environment if the tanks release their contents?	○	○	○	○	○
Pre-conventional	4. What is your level of concern regarding your own personal safety if you choose to stay on-site during the storm?	○	○	○	○	○
Miscellaneous	5. How much faith do you have in your local meteorologist?	○	○	○	○	○
Conventional	7. Would staying to prepare the tanks jeopardize you and your family's safety in the upcoming storm?	○	○	○	○	○

Figure 3.1. Example EPSRI Dilemma with Considerations.

Importance ratings and rankings of relevant considerations were used to quantify respondents' moral reasoning. Specifically, P-scores (principle score) and N2 scores (new principle score 2) were used to benchmark moral development among Kohlberg's three levels of reasoning (pre-conventional, conventional, and post-conventional moral reasoning). These scores originate from earlier validated instruments (DIT and DIT2), and the method of determining these scores from the ratings and rankings within the EPSRI is described in detail elsewhere (Butler, 2018). In addition, variance in rating and ranking of considerations can quantify moral consistency. Neo-Kohlbergian ethicists developed the CDIT score (consistency DIT score) to benchmark respondents' moral reasoning as consolidated or transitional (Rest, Narvaez, Bebeau, et al., 1999).

Instrument Validation

Theory is a set of interrelated constructs that model a phenomena (Creswell, 2009), yet these constructs are often difficult to assess or may not even be observable traits. For example, an individual's feeling empathy or motivation are unobservable constructs. As such, instruments and tests enable researchers to assess the extent an induvial exhibits behaviors or opinions aligning with such constructs. As constructs are not directly observable, questions that theoretically align with the constructs must be developed. Factor analysis is a part of the instrument validation procedure that ensures these questions are valid, or accurately represent the construct, and reliable, or consistently represent the construct (DeVellis, 2016; Meyers et al., 2017b). In the context of the EPSRI, factor analysis checks that the EPSRI's considerations accurately and reliably represent the intended levels of moral reasoning, such as those shown in Figure 3.1. The EPSRI followed the best practices in instrument development, similar to those described in the development of a seminal moral reasoning instrument (DIT2) (Rest, Narvaez, Thoma, et al., 1999) by seeking this evidence of validity and reliability.

Factor analysis is a multivariate testing procedure that searches for correlation among a set of items (Meyers et al., 2017b). As an analogy, the simplest statistical analysis might look for the relationship (or correlation) between two items. Factor analysis scales up this test by looking for the correlation among different groupings of items and reports the correlations from the most robust groupings. These correlated groups of items outline latent constructs (Meyers et al., 2017b). With respect to this study, the items are the ratings of consideration in the EPSRI, and the latent constructs are intended to be the three levels of moral reasoning.

Instrument development follows a four phase process: 1) content generation, 2) content validation, 3) exploratory factor analysis, and 4) confirmatory factor analysis (Hinkin, 1998). Content generation involves writing the ethical dilemmas that are contextualized to the field and considerations that reflect the levels of moral reasoning. Content validation involves reviewing the dilemmas and considerations for appropriateness to the instrument context, which is typically performed by content experts (ethicists, professors, field practitioners, etc.) (Butler et al., 2018; DeVellis, 2016). Exploratory factor analysis (EFA) entails reliability and validity testing of the instrument with the intended sample. EFA may be a cyclic process involving removing EPSRI considerations or dilemmas and retesting the instrument. Upon finding satisfactory reliability and validity in the EFA, development proceeds to confirmatory factor analysis (CFA). CFA repeats the reliability and validity testing used in the EFA to affirm that the reliability and validity is reproducible with a new, independent sample and that the predicted relationships in the data fit well.

Methods

Instrument Validation

Earlier work on the EPSRI included the first two phases of instrument development and two iterations of EFA (Butler, 2018). This prior work obtained evidence of content validity from experts in the process industry and academia, generating seven dilemmas that include thirteen to fifteen considerations per dilemma. The initial two rounds of EFA refined the EPSRI to five dilemmas having twelve to fifteen considerations each; however, these analyses found limited reliability and validity (Butler, 2018). The current study focused on completing a final round of EFA and CFA to obtain sufficient reliability and validity evidence for the EPSRI version with five dilemmas. A summary of these five dilemmas and the EPSRI's development journey are shown in Table 3.1 and Figure 3.2, respectively.

Table 3.1.
Summary of Dilemmas in the EPSRI

Dilemma	Summary of EPSRI dilemma with two ways to respond.
1 (Hose)	While designing a plastic manufacturing facility, you must choose between two metal hoses: a cheap hose that must be replaced frequently or an expensive hose that has a longer life.
2 (Hurricane)	A hurricane is forecasted to flood the chemical plant you work at, which may disable backup generators that sustain temperature controlled chemicals. You may trust in the plant infrastructure to manage flooding or solicit volunteers to manage flooding during the storm.
3 (Leak)	As a new employee, a coworker tells you to ignore a leaking steam valve that you point out. You may report the valve to the manager or ignore the leaking valve.
4 (Gearbox)	As an engineering intern at an oil refinery, a co-worker suggests dismantling a jammed valve gearbox, as this has been done before. A manager may be asked for help, or the gearbox may be dismantled.
5 (Additive)	At a research and development company, a reaction additive you use to boost productivity has recently been reported to be potentially harmful to the environment, but it is not regulated by the EPA. You may search for a replacement or ignore the reports.

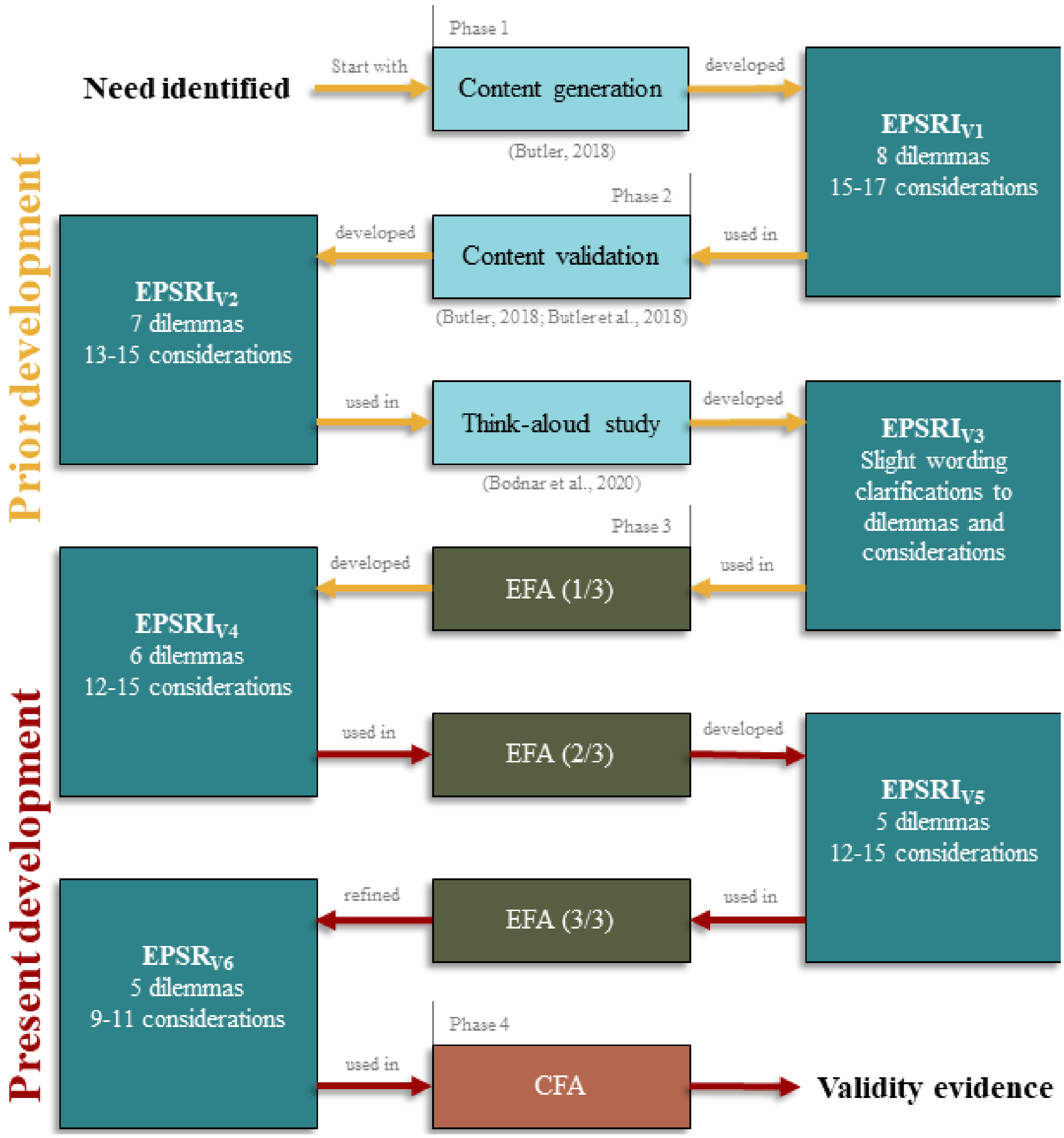


Figure 3.2. Summary of the journey of the EPSRI development while differentiating past and present work.

Data Screening

In all rounds of instrument development, the sample was filtered using three criteria that may indicate responses were unreliable: 1) missing responses, 2) repeat ratings, 3) inconsistence between rate and rank scores. This filtering procedure was derived from those of the DIT and DIT2 (Rest, Narvaez, Thoma, et al., 1999) and has been modified for the context of the EPSRI (Butler, 2018). When taking the EPSRI, respondents read ethical dilemmas, rate the importance of a series of considerations within each dilemma, and rank the top-four considerations to each decision. When too many considerations are unrated or when there are too many missing rankings, the response was removed from the sample. Incomplete responses may be obtained when a respondent decides not to respond to a consideration or quits the test midway through. Excessive missing data makes calculating final scores (p-score, N2 scores, and CDIT scores) inaccurate, and missing responses may lead to the entire data set being unreliable. As such, these responses should be removed. Respondents who left three consideration ratings blank for two dilemmas were purged from the sample (Rest, Narvaez, Thoma, et al., 1999).

Responses may also be removed if too many considerations to a single dilemma are rated of the same importance. When considerations reflect distinct levels of moral reasoning but are repeatedly rated as the same level of importance, this may indicate the respondent unreliably rated all of the considerations similarly to complete the EPSRI quickly. Repeating data was purged where n-1 considerations were rated the same across at least two dilemmas, where n is the number of considerations per dilemma (Rest, Narvaez, Thoma, et al., 1999). As the EPSRI was developed and considerations were removed as the instrument was refined, this threshold was adjusted proportionally with the number of considerations in the dilemma. Moreover, the number of considerations varied between dilemmas, so the threshold also varied between dilemmas within the same version of the EPSRI. For example, Dilemma 1 included 12 considerations, and Dilemma 2 included 15 considerations.

Last, responses may be filtered for inconsistently rating and ranking considerations within a dilemma. For example, a respondent may *rate* some considerations to be of great importance (5) to them but may *rank* others to be in their top-four considerations. These types of responses likely do not capture the moral thinking from the respondent. Logically, considerations of the highest rated importance should also be ranked amongst the most important considerations to a decision. Thus, when these reported considerations are inconsistent, the response was removed for fear of unreliability. Responses may be purged when a respondent's rate-rank score exceeds a threshold; this threshold is a function of the number of considerations in the instrument (the number of opportunities for ratings and rankings to be misaligned). Because the threshold was proportional the number of considerations in a dilemma, the threshold varied for every dilemma across each round of factor analysis where the instrument may have been refined. The specific formula for calculating rate-rank scores and rate-rank thresholds including the example calculations that guided this filter are described in detail elsewhere (Butler, 2018). Table ***3.2*** summarizes the number of responses filtered from each round of instrument development using the specific criteria described in earlier work on the EPSRI (Butler, 2018).

Table 3.2.

Summary of Filtered Data

Sample Description	EFA (3/3)	CFA
Initial sample	372	339
Filtered by missing data criteria	85	24
Filtered by repeating data criteria	8	34
Filtered by rate-rank criteria	16	4
Filtered sample	263	277

Exploratory Factor Analysis

Once the data have been cleaned, factor analysis can begin. Statistical analysis software, such as IBM SPSS, was used for data rotation. Typically, software defaults to determining factor loading based on the sequence that items are included. Data rotation enables the program to start out of sequence with the goal of finding the most robust factor. Because we anticipated correlation between the instrument constructs (the developmental levels of moral reasoning), oblique rotation (DeVellis, 2016; Meyers et al., 2017a; Pett et al., 2011d) was selected.

After data rotation, factor analysis evaluated for eight indicators ((1) correlation matrix, (2) Measure of Sample Adequacy, (3) Kaiser-Meyer-Olkin value, (4) Bartlett's Sphericity test, (5) scree plot, (6) factor structure matrix, (7) inter-item correlation variance, and (8) Cronbach's α coefficient) as described by Pett et al. (2011a, 2011c, 2011b). The correlation matrix (1) checks for appropriateness for factor analysis, ensuring that items have some correlation, but are not identical. The correlation matrix and its determinant are used to find other indicators: Measure of sampling adequacy (MSA), Kaiser-Meyer-Olkin value (KMO), and Bartlett's test. MSA (2), found on the diagonal of the anti-image correlation matrix, is another set of values that should be consulted to ensure that the correlations are adequate for factor analysis. KMO (3) is an indicator that shows if the sample size is appropriate for factor analysis by determining the ratio of participants to the number of items in the analysis. The KMO indicator is equivalent to a priori power analysis. Bartlett's Sphericity test (4) assesses the null hypothesis that relationships exist in the correlation matrix. A statistically significant result would indicate that there is 95% confidence a relationship *doesn't-not* exist. Next, a scree plot (5) suggests how many factors (or latent constructs) might exist in the data. The purpose of the plot is to differentiate definitive factors from scree variance. The factor structure matrix (6) tables the standardized correlation of every item onto a latent factor. These standardized correlations are called factor loadings. Factor loadings can generally be evaluated visually to interpret whether considerations are loading onto the designed constructs. Inter-item correlation variance (7) affirms that there are no two considerations that are *too* similar within a construct. Finally, Cronbach's α coefficient (8) measures how consistently or reliably the considerations load onto the constructs. Generally, a Cronbach's α coefficient greater than 0.70 is considered highly reliable (Cronbach, 1951), yet lower values may still present evidence of reliability (Cortina, 1993; Griethuijsen et al., 2015; Taber, 2018). Factor loadings (validity evidence) and Cronbach's α (reliability) depend on correlations among considerations; thus, whenever a consideration is removed for any reason at any stage in the analysis, the aforementioned indictors must be reassessed as these values are reported while considering a consideration that no longer exists in the instrument.

Factor analysis, in the context of developing the EPSRI, focused on taking considerations that have been written with specific level of moral reasoning in mind and finding statistical relationships among them. When responding to the EPSRI, students ranked a series of “shuffled” considerations based on how important they are to their decisions. Factor analysis intends to find relationships among similar considerations by their shared variance as factor loadings. When a group of considerations load strongly together, the factor may be interpreted as a construct, such as a level of moral reasoning. In this analysis all considerations are included. This includes the meaningless items because the analysis may show if they share a relationship with unintended factors. Figure 3.3 illustrates how an intended meaningless item has loaded onto the interpreted conventional moral reasoning construct. As described later, such a loading may require removal or alteration of the wording of the consideration.

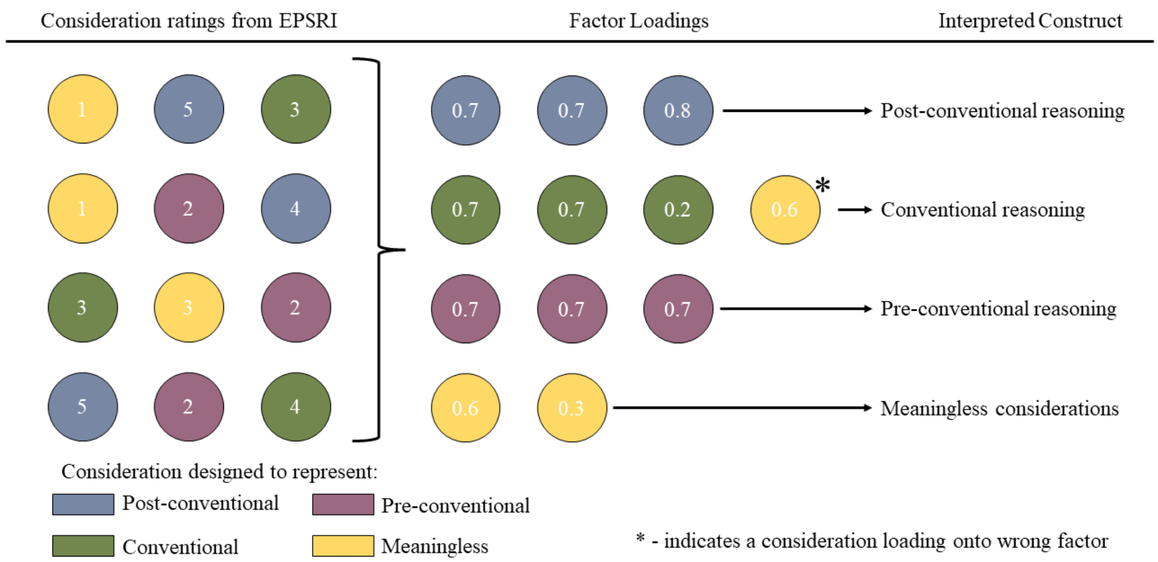


Figure 3.3. Visualization of how EPSRI considerations load onto constructs.

To obtain adequate statistical power for factor analysis, the sample needs to be five to ten times greater than the number of items being analyzed (Hair et al., 1998). The initial version of the EPSRI had almost one hundred items, so a sample of 1000 would have been required. Because of unlikelihood of securing that many senior chemical engineering students in the sample, Butler (2018) made the judgement to perform the EFA by dilemma opposed to across the entire instrument. In this case, a targeted sample of 150 would be appropriate for the maximum fifteen considerations per dilemma. This approach is maintained through subsequent rounds of EFA for analysis consistency.

Confirmatory Factor Analysis

After the EFA, a confirmatory factor analysis (CFA) was performed. While EFA seeks evidence of validity and consistency in a sample, a second sample provides more resounding evidence that the anticipated model is accurate (Gerald Saucier & Goldberg, 1996). In addition, it can promote instrument reliability between samples, suggesting replication. A CFA differs from EFA as it focuses on *confirming* the existence of the hypothesized latent factor structures and relationships among items. Researchers use EFA to identify the optimal number of latent factors and their factor loadings, whereas in CFA, researchers predetermine the relationships and loadings of items into factors, which are informed by the EFA (DeVellis, 2016). There are multiple pathways to perform a CFA because of the flexibility of the analysis (Kline, 2015). Two of which are structural equation modeling (SEM) and repeating principal component analysis (PCA). SEM tests a block diagram model of items with arrows indicating correlations and sources of error (Meyers et al., 2017b). While this method is robust and commonly used (Meyers et al., 2017b), others suggest that testing a provided model against empirical data may provide less evidence of validity than when the data reload onto desired factors on a repeat attempt (DeVellis, 2016; Gerard Saucier & Goldberg, 1996). Following this judgement, the author opted to repeat the PCA procedure over the SEM alternative due to time constraints around this study. As such, the quantitative indicators of validity and reliability are the same between the EFA and CFA.

Study 1 Design

Instrument validation and factor analysis focused on sampling senior chemical engineering students because this is where process safety education is largely taught in chemical engineering programs (CCPS, 2022). A summary of the sampling is shown in Table 3.3. For the first two rounds of EFA, Butler (2018) sampled senior chemical engineering students from three ABET accredited chemical engineering programs

(Programs 1-3), and they obtained a sample size of 223 and 198. These rounds of EFA did not find sufficient evidence of validity and reliability, so some considerations were purged from the instrument. To expand the sample in the third round of exploratory factor analysis, data was collected from seven new ABET accredited chemical engineering programs (Programs 4-10), obtaining a filtered sample of 263 senior chemical engineering students. Upon obtaining adequate reliability and validity of the intended constructs in the EFA, the CFA was performed by sampling from five programs (Programs 4-8) to obtain 275 responses. Programs selected for each of these rounds of analysis were selected as part of the research network of study collaborators overseeing the larger research project.

Table 3.3.
Summary of Sample across Instrument Development

Round of development	Sample size	Institutions
EFA 1/3 (spring 2018) *	210	1-3
EFA 2/3 (fall 2018) *	198	1-3
EFA 3/3 (spring 2019)	263	4-10
CFA (spring 2020)	275	4-8

* - indicates rounds of instrument development performed prior to this work

Results and Discussion

Third Exploratory Factor Analysis

This present work began with performing the third round of EFA. As a result of the second round of EFA (Butler, 2018), the EPSRI was refined into its fifth version. The number of dilemmas and considerations per dilemma for this version of the EPSRI is summarized in Table 3.4.

Table 3.4
Number of Considerations Designed into the EPSRI (version 5) at the End of the Second EFA to be Used in the Beginning of the Third EFA

	Number of considerations				
Dilemma	Pre-conventional	Conventional	Post-conventional	M-item	Total
1 (Hose)	4	3	3	2	12
2 (Hurricane)	4	4	4	3	15
3 (Leak)	4	4	3	1	12
4 (Gearbox)	3	3	4	3	13
5 (Additive)	4	4	3	2	13

When evaluating the first ethical dilemma, in which participants are asked to select a hose for a manufacturing process, one post-conventional consideration was removed because it was loading strongly onto a factor along with three other conventional considerations. This loading suggests that students were interpreting this consideration to similarly reflect conventional level reasoning, which was not the intended case. The consideration asked, "(7) Do you believe there is a level of risk that is acceptable considering the savings of Option A?" After removing this consideration, the remaining four factor analysis indicators (MSA, KMO, Bartlett's test, and scree plots) supported performing a factor analysis with four distinct factors with considerations consistent with their intended design. A summary of this dilemma's factor loadings and factor reliabilities are shown in Table 3.5. Note, the removed consideration is not displayed in this table as its inclusion in the analysis would impact the other loadings in the analysis.

Table 3.5.
Summary of Dilemma 1's Factor Loadings and Reliabilities from the Third Round of EFA

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(1-1) How important is it to you to maximize your salary with a bonus?	0.64	0.51
	(1-3) Would you gain personal satisfaction of "doing a good job" if you choose one option over the other?	0.36	
	(1-8) Would your fellow employees have a more positive opinion of you if you chose one option or the other?	0.68	
	(1-11) Would you earn additional recognition and career opportunities by keeping costs low?	0.78	
Conventional	(1-2) Are you concerned that selecting Option A might require more of your employees' time?	0.49	0.42
	(1-5) How much risk to the surrounding community is associated with replacement of hoses for each option?	0.76	
	(1-10) Would choosing Option B benefit the company?	0.58	

	Consideration	Factor loading	Reliability, Cronbach's α
Post-conventional	(1-6) Would process reliability be affected by hose choice?	0.81	0.62
	(1-7) Is it ever a good idea to rely on active measures (in this case, employee maintenance) rather than inherently safe design (in this case, material of construction)?	0.79	
M-Item	(1-4) Do you think that management will cancel work due to weather tomorrow?	-0.83	0.59
	(1-9) Do the polymer linings for each option match the color of the pipe?	-0.84	

When evaluating the second ethical dilemma, where participants are asked to decide how to prepare for a hurricane, four considerations were purged from Dilemma 2. Of those four, two considerations were removed for having MSA values below the desired threshold of 0.540 (Hair et al., 1998; Meyers et al., 2017b). Following the removal of these first two considerations, the next two considerations were removed for loading onto unintended factors. Consideration 2-13 (M-Item) loaded strongly onto the pre-conventional factor. When this consideration was removed, it left a single M-Item in this dilemma. Any single consideration cannot represent a latent factor, so when the author repeated the procedure after the removing the aforementioned considerations, the single remaining M-Item was also not included. After its removal, Consideration 2-3 (conventional) loaded onto both the pre-conventional and post-conventional factors, so this consideration was also removed from the dilemma. A summary of the considerations removed from this dilemma are shown in Table 3.6. The remaining considerations were found to be appropriate for factor analysis, and each loaded strongly to their intended factor. A summary of this dilemma's factor loadings and factor reliabilities are shown in Table 3.7.

Table 3.6.
Considerations in Sequence of Removal from Dilemma 2 during the Third Round of EFA

Consideration	Reason for removal
(Pre-conventional Consideration 2-2) Are you concerned about your job security if the plant is damaged or destroyed?	The MSA value of 0.44 suggested its correlations with other considerations were not appropriate for factor analysis.
(Pre-conventional Consideration 2-14) Would working at the plant during the storm impact your ability to protect your home and belongings from damage?	The MSA value of 0.49 suggested its correlations with other considerations were not appropriate for factor analysis.

Consideration	Reason for removal
(Conventional Consideration 2-3) Are you concerned about your coworkers' safety if they stay at the plant during the hurricane?	Dual loaded onto opposing constructs. Pre-conventional loading: 0.447 Post-conventional loading: 0.583
(Conventional Consideration 2-13) What is the potential for wind damage to the plant?	The MSA value of 0.44 suggested its correlations with other considerations were not appropriate for factor analysis.

Table 3.7.
Summary of Dilemma 2's Factor Loadings and Reliabilities

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(2-6) What is your level of concern regarding your own personal safety if you choose to stay on-site during the storm?	0.76	0.49
	(2-9) Would staying to prepare the tanks jeopardize you and your family's safety in the upcoming storm?	0.77	

	Consideration	Factor loading	Reliability, Cronbach's α
Conventional	(2-5) What is your level of comfort in soliciting volunteers to stay on-site during what may be a life-threatening situation?	0.74	0.50
	(2-8) What is your confidence that you and/or your team will be able to keep the generators functioning under the storm conditions?	0.70	
	(2-10) How would your company be perceived if no employees were on site during the storm and an accident occurred?	0.66	
Post-conventional	(2-1) What is the potential for negative impact to the environment if the tanks release their contents?	0.67	0.69
	(2-4) Is there the potential for the exploding tanks to damage the surrounding neighborhood and infrastructure adjacent to the plant?	0.71	
	(2-12) What is the potential for negative health effects on residents who live in the areas surrounding the plant if the tanks release their contents?	0.87	

	Consideration	Factor loading	Reliability, Cronbach's α
Post-conventional	(2-15) Is the plant located in an area where dangerous impacts to the surrounding community are likely?	0.44	
M-Item	(2-7) How much faith do you have in your local meteorologist?	1.0	-

When evaluating the third ethical dilemma regarding a leaking steam valve, one consideration was removed from Dilemma 3. All of the eight factor analysis indicators supported performing an EFA, and the factor loadings were strong and grouped as intended. However, when evaluating the loading reliability for the factor with the four conventional considerations, Consideration 3-7 was removed to improve the reliability from $\alpha = 0.47$ to $\alpha = 0.52$. Consideration 3-7 asked respondents, "Would there be any negative impacts on your coworker if you report the leak?" A summary of this dilemma's factor loadings and factor reliabilities are shown in Table 3.8.

Table 3.8.
Summary of Dilemma 3's Factor Loadings and Reliabilities

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(3-4) Will the leaking valve be a common nuisance for you, or is it located in a part of the plant you will seldom visit?	0.49	0.61
	(3-8) Will you face negative repercussions from reporting the leak?	0.70	
	(3-10) How much time or effort would it take you to have the valve inspected?	0.60	
	(3-11) Would you be looked at as a "worrier" if you report the leak?	0.76	
Conventional	(3-1) How often is maintenance performed on the equipment in the plant?	0.66	0.52
	(3-5) How often is the valve used?	0.20	
	(3-8) What other people or equipment may be exposed to the steam leak?	0.709	

	Consideration	Factor loading	Reliability, Cronbach's α
Post-conventional	(3-2) If you ignore the leaky valve, are you contributing to a negative engineering safety culture overall?	-0.33	0.48
	(3-6) Is it ever acceptable to not report a potential safety hazard?	0.86	
	(3-12) Can any safety hazard, regardless of how minor, be dismissed as simply an "annoyance"?	0.82	
M-Item	(3-3) How big of a bucket is available to collect the leaking material from the valve?	1.0	-

When evaluating the fourth ethical dilemma regarding dismantling a valve gearbox at an oil refinery, a pre-conventional consideration was removed for a low MSA value. Upon reperforming the analysis after its removal, three more considerations were subsequently removed for loading onto factors in addition to their intended factors. A summary of the considerations removed are shown in Table 3.9. A summary of this dilemma's factor loadings and factor reliabilities are shown in Table 3.10.

Table 3.9.
Considerations Removed from Dilemma 4 During the Third Round of EFA

Consideration	Reason for removal
(Pre-conventional Consideration 4-12) Are you concerned about your job security if your supervisor becomes aware that you opened the valve manually?	The MSA value of 0.46 suggested its correlations with other considerations were not appropriate for factor analysis.
(Conventional Consideration 4-8) Are you comfortable deviating from the company's standard operating procedures?	Removed to improve reliability of conventional considerations from $\alpha = 0.15$ to $\alpha = 0.29$.
(M-Item 4-6) Does the weather forecast predict cloudy weather today?	Removed to improve reliability of meaningless considerations from $\alpha = 0.55$ to $\alpha = 0.75$.
(Post-conventional Consideration 4-13) Is it right to regularly manually override equipment?	Removed to improve reliability of post-conventional considerations from $\alpha = 0.63$ to $\alpha = 0.69$.

Table 3.10.
Summary of Dilemma 4's Factor Loadings and Reliabilities

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(4-2) Would your co-worker lose confidence in your abilities if you asked for assistance?	0.91	0.80
	(4-3) Would the engineering supervisor be irritated with you if you asked for help?	0.89	
Conventional	(4-5) Would leaving the valve as-is cause a loss of production?	0.77	0.29
	(4-9) Would a hazardous chemical leak impact the company's image?	0.58	
Post-conventional	(4-1) Are there any health risks associated with a hazardous chemical leak that could impact the local community?	0.80	0.75
	(4-6) What is the possibility of a larger issue, such as an explosion or fire, if a hazardous chemical were to leak from the valve?	0.86	
	(4-7) What impact would a hazardous chemical leak have on the environment?	0.84	
M-Item	(4-4) How many bolts will need to be undone?	0.83	0.69
	(4-8) What tools do I need to unbolt the valve?	0.81	

When evaluating the fifth ethical dilemma regarding selecting a potentially environmentally harmful process alternative, one M-Item was removed for a low MSA value. Upon its removal, the factor analysis indicators supported using the data for factor analysis, and the considerations all loaded strongly onto their intended factors. One post-conventional consideration was later removed to improve the reliability of the factor. A summary of the considerations removed are shown in Table 3.11. A summary of this dilemma's factor loadings and factor reliabilities are shown in Table 3.12.

Table 3.11.
Considerations Removed from Dilemma 5 during the Third Round of EFA

Consideration	Reason for removal
(M-Item 5-5) What types of products is the additive used to make?	MSA=0.460
(Post-conventional Consideration 5-12) Is it right to discharge a chemical to the environment that is suspected to be hazardous?	Cronbach's α rose from 0.485 to 0.611

Table 3.12.
Summary of Dilemma 5's Factor Loadings and Reliabilities

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(5-2) Are you concerned that your yearly bonus will be impacted if your company discontinues the use of this chemical?	0.42	0.49
	(5-5) What is the difficulty and personal time investment it will take for you to find a replacement additive?	0.46	
	(5-8) Are you concerned about your health since you are constantly exposed to the additive?	0.46	
	(5-10) Are you concerned about your job security if you should fail to find an appropriate alternative?	0.72	
Conventional	(5-1) What is the potential for lost production if you discontinue the additive without finding a suitable replacement?	0.63	0.66
	(5-3) What is the potential impact on your coworkers' jobs if the additive is banned and you didn't look for an alternative?	0.25	

	Consideration	Factor loading	Reliability, Cronbach's α
Conventional	(5-6) How important to the company is it that the government agency in charge of environmental regulations has not issued any ruling on the continued use of this additive?	0.84	
	(5-9) Is there any additional time or money it would cost your company to replace the additive?	0.78	
Post-conventional	(5-4) Are you concerned that discharging the chemical could impact the local community?	0.78	0.61
	(5-7) What is the potential for negative environmental and human consequences if the additive is eventually proven to be dangerous?	0.78	
M-Item	(5-11) Are you concerned that the replacement might have a name that is difficult to pronounce?	1.0	-

In summary, this round of EFA identified and removed considerations whose correlations were inappropriate for factor analysis, which loaded onto unintended factors, or that could be removed to improve the reliability of the construct. Upon these changes, the remaining considerations found strong evidence of validity according to their factor loadings; when factors load so strongly onto intended factor, it leaves little question as to what construct those considerations represent (Kline, 2015). Next, a number of the extracted factors exhibited limited reliability using Cronbach's α coefficient. Within each dilemma there are few considerations designed to represent a level of moral reasoning; this inherently will reduce the maximum obtainable reliability (Griethuijsen et al., 2015; Taber, 2018). In addition, some variance in reliability among the extracted factors is to be expected in instrument as it is unlikely that each factor would yield identical reliability scores. However, these values may speak more to the limited sample size than the actual reliability of the responses (Griethuijsen et al., 2015; Taber, 2018). It is by these results that the author argues that the EPSRI has found sufficient validity evidence, warranting a confirmatory factor analysis.

Confirmatory Factor Analysis

The final round of EFA removed a few considerations from the previous version of the EPSRI, creating the sixth version. The CFA further assesses this version of the EPSRI for evidence of validity and reliability. A summary of this version of the instrument is shown in Table 3.13. As this phase of factor analysis is purely confirming the previous findings, no further considerations were revised or removed.

Table 3.13.
Number of Considerations Designed into the EPSRI (Version 6) at the End of EFA 3 to be Used in the CFA

Dilemma	Number of considerations				
	Pre-conventional	Conventional	Post-conventional	M-item	Total
1 (Hose)	4	3	2	2	11
2 (Hurricane)	2	3	4	1	10
3 (Leak)	4	3	3	1	11
4 (Gearbox)	2	2	3	2	9
5 (Additive)	4	4	2	1	11

In each of the analyzed dilemmas, the eight factor analysis indicators were all within desired thresholds. Considerations' factor loadings and constructs' reliabilities by dilemma are shown in Table 3.14, Table 3.15, Table 3.16, Table 3.17, and Table 3.18. These tables show that the factor loadings and reliabilities are quite consistent between the third EFA and CFA. The repeatedly strong factor loadings provide evidence of discriminant validity. While some of the reported Cronbach's α values are on the lower end of the reliability spectrum, the consistency of factor loadings and Cronbach's α values support repeatability in the EPSRI results.

Table 3.14.
Summary of Dilemma 1's Factor Loadings and Reliabilities from the CFA

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(1-1) How important is it to you to maximize your salary with a bonus?	0.62	0.55
	(1-3) Would you gain personal satisfaction of "doing a good job" if you choose one option over the other?	0.68	
	(1-8) Would your fellow employees have a more positive opinion of you if you chose one option or the other?	0.59	
	(1-11) Would you earn additional recognition and career opportunities by keeping costs low?	0.69	
Conventional	(1-2) Are you concerned that selecting Option A might require more of your employees' time?	0.23	0.36
	(1-5) How much risk to the surrounding community is associated with replacement of hoses for each option?	0.72	
	(1-10) Would choosing Option B benefit the company?	0.80	

	Consideration	Factor loading	Reliability, Cronbach's α
Post-conventional	(1-6) Would process reliability be affected by hose choice?	0.86	0.65
	(1-7) Is it ever a good idea to rely on active measures (in this case, employee maintenance) rather than inherently safe design (in this case, material of construction)?	0.83	
M-Item	(1-4) Do you think that management will cancel work due to weather tomorrow?	-0.80	0.55
	(1-9) Do the polymer linings for each option match the color of the pipe?	-0.81	

Table 3.15.

Summary of Dilemma 2's Factor Loadings and Reliabilities from the CFA

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(2-4) What is your level of concern regarding your own personal safety if you choose to stay on-site during the storm?	0.77	0.47
	(2-7) Would staying to prepare the tanks jeopardize you and your family's safety in the upcoming storm?	0.77	
Conventional	(2-3) What is your level of comfort in soliciting volunteers to stay on-site during what may be a life-threatening situation?	0.80	0.38
	(2-6) What is your confidence that you and/or your team will be able to keep the generators functioning under the storm conditions?	0.81	
	(2-8) How would your company be perceived if no employees were on site during the storm and an accident occurred?	0.24	
Post-conventional	(2-1) What is the potential for negative impact to the environment if the tanks release their contents?	0.70	0.68

	Consideration	Factor loading	Reliability, Cronbach's α
Post-conventional	(2-2) Is there the potential for the exploding tanks to damage the surrounding neighborhood and infrastructure adjacent to the plant?	0.70	
	(2-10) What is the potential for negative health effects on residents who live in the areas surrounding the plant if the tanks release their contents?	0.79	
	(2-11) Is the plant located in an area where dangerous impacts to the surrounding community are likely?	0.65	
M-Item	(2-5) How much faith do you have in your local meteorologist?	1.0	-

Table 3.16.
Summary of Dilemma 3's Factor Loadings and Reliabilities from the CFA

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(3-4) Will the leaking valve be a common nuisance for you, or is it located in a part of the plant you will seldom visit?	0.64	0.64
	(3-7) Will you face negative repercussions from reporting the leak?	0.72	
	(3-9) How much time or effort would it take you to have the valve inspected?	0.58	
	(3-10) Would you be looked at as a "worrier" if you report the leak?	0.78	
Conventional	(3-1) How often is maintenance performed on the equipment in the plant?	0.77	0.64
	(3-5) How often is the valve used?	0.73	
	(3-8) What other people or equipment may be exposed to the steam leak?	0.76	

	Consideration	Factor loading	Reliability, Cronbach's α
Post-conventional	(3-2) If you ignore the leaky valve, are you contributing to a negative engineering safety culture overall?	-0.28	0.49
	(3-6) Is it ever acceptable to not report a potential safety hazard?	0.86	
	(3-11) Can any safety hazard, regardless of how minor, be dismissed as simply an "annoyance"?	0.90	
M-Item	(3-3) How big of a bucket is available to collect the leaking material from the valve?	1.0	-

Table 3.17.
Summary of Dilemma 4's Factor Loadings and Reliabilities from the CFA

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(4-2) Would your co-worker lose confidence in your abilities if you asked for assistance?	0.90	0.80
	(4-3) Would the engineering supervisor be irritated with you if you asked for help?	0.91	
Conventional	(4-5) Would leaving the valve as-is cause a loss of production?	0.57	0.21
	(4-9) Would a hazardous chemical leak impact the company's image?	0.86	
Post-conventional	(4-1) Are there any health risks associated with a hazardous chemical leak that could impact the local community?	0.82	0.73
	(4-6) What is the possibility of a larger issue, such as an explosion or fire, if a hazardous chemical were to leak from the valve?	0.81	
	(4-7) What impact would a hazardous chemical leak have on the environment?	0.80	
M-Item	(4-4) How many bolts will need to be undone?	0.92	0.83
	(4-8) What tools do I need to unbolt the valve?	0.91	

Table 3.18.
Summary of Dilemma 5's Factor Loadings and Reliabilities from the CFA

	Consideration	Factor loading	Reliability, Cronbach's α
Pre-conventional	(5-2) Are you concerned that your yearly bonus will be impacted if your company discontinues the use of this chemical?	-0.06	0.56
	(5-5) What is the difficulty and personal time investment it will take for you to find a replacement additive?	0.41	
	(5-8) Are you concerned about your health since you are constantly exposed to the additive?	0.242	
	(5-10) Are you concerned about your job security if you should fail to find an appropriate alternative?	0.727	
Conventional	(5-1) What is the potential for lost production if you discontinue the additive without finding a suitable replacement?	0.57	0.51
	(5-3) What is the potential impact on your coworkers' jobs if the additive is banned and you didn't look for an alternative?	0.23	

	Consideration	Factor loading	Reliability, Cronbach's α
Conventional	(5-6) How important to the company is it that the government agency in charge of environmental regulations has not issued any ruling on the continued use of this additive?	0.57	
	(5-9) Is there any additional time or money it would cost your company to replace the additive?	0.71	
Post-conventional	(5-4) Are you concerned that discharging the chemical could impact the local community?	0.76	0.64
	(5-7) What is the potential for negative environmental and human consequences if the additive is eventually proven to be dangerous?	0.80	
M-Item	(5-11) Are you concerned that the replacement might have a name that is difficult to pronounce?	1.0	-

Reliability

Between the EFA and CFA, the reliability results using Cronbach's α from the EFA and CFA generally ranged from ~0.50 (sufficient) to ~0.80 (robust) (Taber, 2018). Two factors may have promoted this outcome: (1) performing the factor analysis on each individual dilemma and (2) removal of considerations. Prior factor analysis on the EPSRI opted to analyze the instrument at each ethical dilemma instead of across the whole instrument; this decision was motivated by the need for statistical power found through the ratio between the sample size and the number of considerations in the instrument (Butler, 2018). When calculating Cronbach's α to determine reliability, the calculation considers the number of items in the analysis (Taber, 2018). When performed at the dilemma level, there are five to seven times fewer considerations applied to this analysis, which inherently imposes an upper limit on the obtainable reliability scores (Griethuijsen et al., 2015). Multiple reports question Cronbach's α analytical sensitivity to the number of items in a construct that can "fix" highly reliable outcomes (Cortina, 1993; Taber, 2018). Also, while performing the EFA, the author used the Principal Component Analysis (PCA) method for consistency with prior work (Butler, 2018). PCA refines instruments to identify which items best represent the intended constructs and which can be removed (DeVellis, 2016). As a result, few considerations remained in the EPSRI when moving into the CFA, which may have also contributed to the obtained reliability scores even when the factor loadings remained the same (DeVellis, 2016).

While the range of reliability results were on the lower end of what is considered acceptable ($\alpha > 0.70$) (Cronbach, 1951), the consistent results from the factor loadings between rounds of factor analysis (EFA and CFA) support the intended constructs. Kline (2015) suggests that such repeated loadings leave little question of the intended constructs. As such, the author evaluates the EPSRI to provide adequate evidence of reliability despite controversial Cronbach' α results.

Limitations

The sample used in the CFA was collected early in the 2020 spring semester, which coincided with the beginning of the COVID-19 pandemic. It is possible that public messaging around community safety and PPE may have influenced EPSRI responses. Although the factor loadings between the EFA and CFA were consistent, the responses used in the CFA may not be representative of other responses obtained post-pandemic because of the lack of active safety messaging.

Also, the samples used in both the EFA and the CFA were limited in size. Factor analysis was pursued on the dilemma level instead of the instrument level to mitigate the effects of low statistical power (Butler, 2018). However, this methodological decision made trade-offs in the maximum obtainable reliability score (Kline, 2015). As such, researchers and instructors should consider evaluating the responses to the EPSRI in their data for internal consistency, despite this study finding sufficient evidence of validity and reliability.

Conclusions

Investigative case studies on process incidents have shown how engineers' decision making can create pathways to plant failures despite preventative barriers (CSB, 2015, 2018, 2019b; Ness, 2015). Failures may occur from decision making biases, such as behavioral forecasting or ethical fading (Bazerman & Tenbrunsel, 2011; Osberg & Shrauger, 1986; A. Tenbrunsel et al., 2010). Experts suggest that these failures may be mitigated through decision making practice environments and education (Barab & Duffy, 2012; Nicaise et al., 2000; Radinsky et al., 2001), yet no form of assessment has existed that accounts for the nuances of ethical dilemmas in the chemical process industry. As such, this study completed the development of the Engineering Process Safety Research Instrument (EPSRI).

The final version of the EPSRI specifically assesses the moral reasoning of respondents through five ethical dilemmas, each with nine to eleven considerations relevant to the dilemma. Development took place through three rounds of EFA (one reported in this study) and one round of CFA. The combination of these analyses provided sufficient evidence that the EPSRI can assess the moral reasoning of respondents. Testing the data with Cronbach's α did highlight that there may be limitations to the reliability of the instrument (Cronbach, 1951), but these values may be attributed to the limited sample sizes obtained (Pearson & Mundfrom, 2010). However, the strong, repeatedly obtained correlations among the considerations support the instrument reflects the intended levels of moral reasoning.

In completing this instrument development, the author encourages process safety instructors to consider incorporating decision-making education because of its potential impact for the chemical process industry. Furthermore, the author encourages instructors who choose to assess their students' moral reasoning with the EPSRI to understand its limitations and encourages them to complete their own reliability assessment with their selected sample population.

Chapter 4

Efficacy of Digital Immersive Educational Games

Overview

The objective of this second study is to assess the impact of a digital immersive game on the moral reasoning of senior chemical engineering students. The literature review on ethical decision-making biases suggests that engineers are likely to predict they will behave more ethically than their actual actions. Training through an immersive digital game holds the potential for engineering students to build awareness of their behavioral ethics, correcting errors in their behavioral forecasting (Bairaktarova & Woodcock, 2017; Clancy, 2020b; Osberg & Shrauger, 1986). This study is guided by the overarching research question *how does a digital process safety game influence the ethical decision making of chemical engineering students?*

Specifically, this study evaluates students' moral reasoning using Kohlberg's theory of moral development. This study first pairs senior chemical engineering students between two modes of assessment: the Engineering Process Safety Research Instrument (EPSRI) and a digital process safety decision making game, Contents Under Pressure. Students' responses to moral reasoning reflection prompts are compared between the two modes. In addition, the impact of Contents Under Pressure on the students' moral reasoning is evaluated in a retrospective analysis between a comparison cohort (standard process safety curriculum) and an intervention cohort (curriculum with game added). Moral reasoning is evaluated with a combination of quantitative N2 and CDIT scores extracted from responses to the EPSRI. The results suggest that students' responses withing Contents Under Pressure are authentic unlike responses within the EPSRI because the responses better align with expectations from literature. Responses within the EPSRI were likely inflated due to behavioral forecasting. Moreover, Contents under Pressure gameplay is strongly correlated with students becoming prepared to make advancements in their moral reasoning.

Introduction

Engineering Process Safety Research Instrument

The EPSRI is a moral reasoning instrument specific to process safety dilemmas, as described in Chapter 3. The instrument presents respondents with five ethical dilemmas, each with a series of possible considerations to the respondents' decision. These considerations were written to align with the three levels of Kohlberg's theory of moral development. Examples of these considerations from the EPSRI are given in Table 4.1.

Table 4.1.
Example Considerations from the EPSRI.

Level of moral reasoning	Example EPSRI consideration
Pre-conventional (Consideration 2-4)	What is your level of concern regarding your own personal safety if you choose to stay on-site during the storm?
Conventional (Consideration 2-3)	What is your level of comfort in soliciting volunteers to stay on-site during what may be a life-threatening situation?
Post-conventional (Consideration 2-10)	What is the potential for negative health effects on residents who live in the areas surrounding the plant if the tanks release their contents?

A respondent's moral reasoning may be determined based on how they rate and rank these considerations. Quantitative methods described in the work around the DIT2 enable extracting N2 scores, CDIT scores, and predominant level of moral reasoning (Rest, Thoma, & Edwards, 1997). The N2 is a scale variable that indicates the degree of ethicality a respondent expresses; a higher score suggests higher levels of moral reasoning. The CDIT score describes how consistently a respondent reasons at one of three levels of moral reasoning. A score greater than 15.705 indicates consolidated reasoning that primarily leverages a single level, and a score less than 15.705 indicates transitional reasoning that leverages multiple levels (Rest, Narvaez, Bebeau, et al., 1999). Example calculations for these scores are given in Butler et al. (2019). Moral consolidation or transition per level of moral reasoning create six moral *schema*.

Contents Under Pressure

The digital process safety game evaluated in this study is Contents Under Pressure. The premise of the game is to engage students in a narrative as newly hired senior plant manager. As managers, students oversee a team of engineering operators, respond to a safety supervisor and the plant's head chief, and stay dedicated to family members. The cast of characters are shown in Figure 4.1.

Figure 4.1. Cast of characters students interact with in Contents Under Pressure.

Throughout these interactions, students must balance and maintain an array of game metrics relevant to the plant's operations. These metrics include time in the day, overall safety, reputation among colleagues, and plant production, which are shown using emoticons on the top of the game screen in Figure 4.2. Students respond to in game dilemmas by making binary decisions ranging from the mundane (going out to lunch and diffusing coworker squabbles) to the critical (responding to employee injuries and rescuing employees from a major tropical storm). Students may respond in whichever way they please, but every decision has implications on some metric in the game. Implications on game metrics are hinted at as the emoticons flash when hovering the mouse over the binary answer, yet the degree and direction of impact are not specified with the flashing.

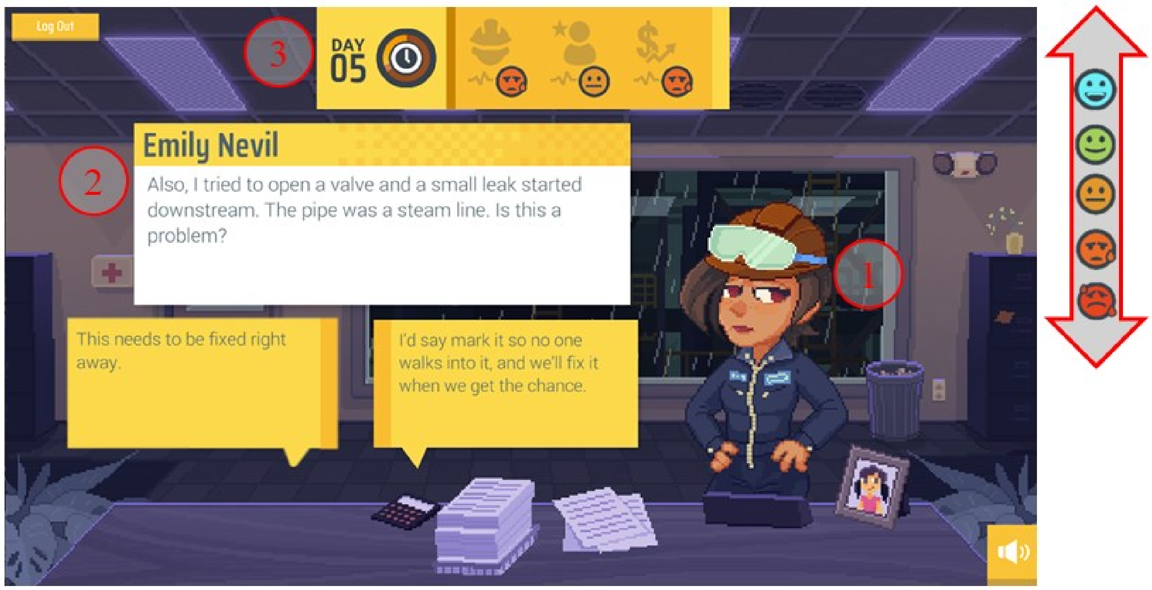


Figure 4.2. Screenshot of Contents Under Pressure gameplay with callouts. 1) Characters in the game appear on the right side of the screen with unique voices and dynamic facial expressions. 2) Characters pose questions and problems, and hovering the mouse over the cards reveal the binary responses. 3) Students must balance game metrics shown at the top of the screen with a depleting clock and three emoticons, which change colors and expression based on their state.

After certain decisions, the players are asked to rate the importance of a single consideration to their decision on a five-point scale through a reflection prompt. For example, after deciding whether to give an employee an extension to writing a report, a reflection prompt may ask the student whether they considered that the report was for their boss. Similar to the EPSRI, these reflection prompts are designed to represent one of the three levels of moral reasoning from Kohlberg's theory. Examples of reflection prompts for each level of moral reasoning are given in Table 4.2. Gameplay spans 15 consecutive calendar days, where each day includes up to 21 decisions taking an average of 13-minutes per session. A video example of gameplay is shown elsewhere (Burkey et al., 2020).

Table 4.2.
Example Contents Under Pressure Reflection Prompts

Level of moral reasoning	Example CUP consideration
Pre-conventional	Balancing your personal life with your work life can be difficult. How relevant is this to your decision?
Conventional	Giving a staff member a raise could increase their dedication to work at the plant. How relevant is this to your decision?
Post-conventional	Community impacts from plant leaks can be devastating. How relevant is this to your decision?

Study 2 Research Questions

The work in Study 2 answers three research sub-questions as building blocks to answer one overarching question, *how does a digital process safety game influence the ethical decision making of chemical engineering students?* These sub-questions are framed around Kohlberg's theory of moral development and are as follows:

4.1 What differences exist between students' expressed moral reasoning when taking surveys and when interacting with a more immersive digital game?

4.2 How does participation in a digital immersive process safety environment influence the moral reasoning of senior chemical engineering students?

4.3 What differences exist in the moral schema of senior chemical engineering students once exposed to a digital immersive process safety environment in comparison to standard process safety instruction?

Broadly, this study found that students may have inflated their expressed moral development due to a phenomenon known as behavioral forecasting. As a result, responses to similar moral reasoning considerations were rated higher in the context of the EPSRI than in Contents Under Pressure. This may be due to the immersive nature of the game diminishing the effects of behavioral forecasting. In the retrospective analysis, N2 score extractions were similar between the intervention and comparison cohorts. However, the intervention cohort did have significant changes to their moral schema and became significantly more morally transitional. These findings suggest that Contents Under Pressure may be preparing students to advance their moral reasoning in the future. Future steps in assessing moral reasoning and teaching process safety decision making should include longitudinal work, as it is presently unclear the effects of an immersive game on students beyond graduation.

Methods

Study 2 Design

This study sampled 298 students from three ABET accredited chemical engineering programs in the fall semesters of 2018 and 2019. Programs were selected as part of the research network of study collaborators. Sample breakdown for each research sub-question and year of sampling is shown in Figure 4.3.

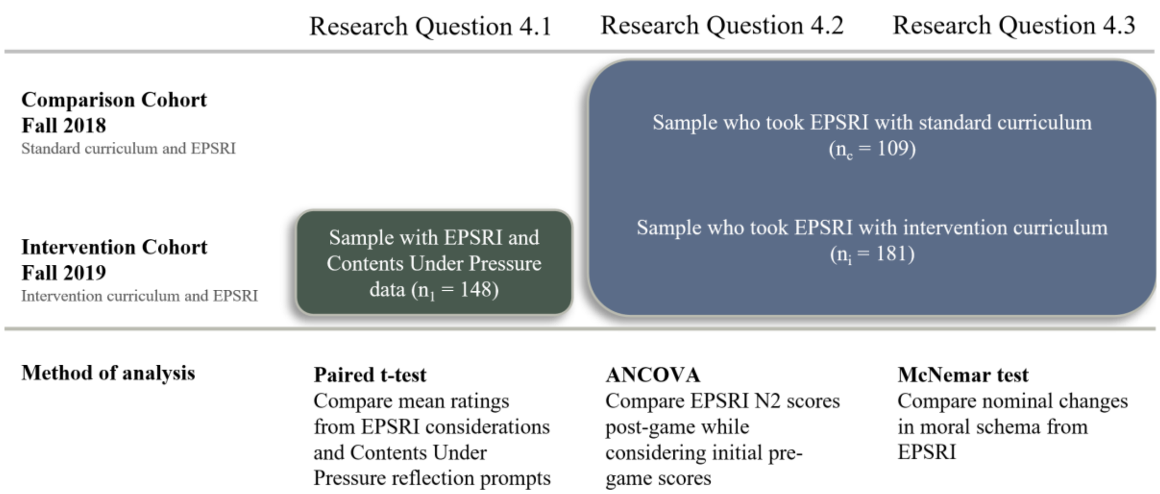


Figure 4.3. Visual of curriculum format by year and sample used to answer each research sub-question.

The 2018 sample has students respond to the EPSRI at the beginning (pre-curriculum) and end (post-curriculum) of the semester. The 2019 sample has students respond to the EPSRI towards the beginning of the semester (pre-intervention), play Contents Under Pressure mid-semester, and respond to the EPSRI again towards the end of the semester (post-intervention). Pairing Contents Under Pressure gameplay and pre-intervention EPSRI responses enables comparing differences in moral reasoning between the two modes of assessment as a way of answering Research Sub-question 4.1. Comparing the EPSRI responses from the 2018 standard curriculum and the 2019 intervention curriculum provides insight on the impact of the digital game intervention. In this analysis, both the 2018 and 2019 EPSRI responses are paired across the semester: pre-curriculum with post-curriculum and pre-intervention with post-intervention. These data were filtered by "missingness" because decisions cannot be statistically interpolated (Osborne, 2012b). This study requires students' initial reaction to ethical dilemmas, as repeated gameplay may confound students' responses by permitting them to reflect on the implications of their decisions. Thus, any student who played a day of Contents Under Pressure more than once was filtered from the sample for the first research sub-question. Note that the first research sub-question (n_1) has a smaller sample size than the second and third research sub-questions for the same year (n_i) because the first research sub-question needed to pair the EPSRI and Contents Under Pressure data. While Contents Under Pressure was a required component of the intervention curriculum, it is possible students did not elect to have their data collected through the game browser or they were filtered for missing reflections. In contrast, the second and third research sub-questions were answered using data solely from the EPSRI, which enabled this sample (n_i) to be larger than the first research sub-question sample (n_1) due to fewer data filters.

In both of these courses, the EPSRI was administered digitally as classwork at both the beginning and end of the course. The author filtered EPSRI data by "unreliable or bogus data" (Butler, 2018; Rest, Narvaez, Thoma, et al., 1999). In the 2019 sample, Contents Under Pressure was implemented approximately four weeks after and four weeks before the EPSRI to mitigate bias between the modes of assessment. Contents Under Pressure gameplay was assigned as homework to be completed over the course of four weeks for completion credit. While Contents Under Pressure's narrative only spans 15-days, additional time was given to students to ensure completion.

Research Question 4.1

The first research sub-question asked, "*what differences exist between students' expressed moral reasoning when taking surveys and when interacting with a more immersive digital game?*" Specifically, it sought to understand the differences in moral reasoning between responses to the EPSRI and Contents Under Pressure. The author answers this research sub-question by comparing how students rate the importance of considerations in the EPSRI and in Contents Under Pressure's reflection prompts. To promote the validity of the comparison between these two modes of assessment, the author only used data from a single EPSRI dilemma because that dilemma reflected the narrative extended through Contents Under Pressure.

Given that EPSRI and Contents Under Pressure considerations were designed to reflect the three levels of moral reasoning, the author took each respondent's mean rating for each of the three levels of moral reasoning for both of the modes of assessment. This procedure created six mean ratings for each respondent (three for each of the EPSRI and Contents Under Pressure, where each align with the three levels of moral reasoning). To begin the analysis, the levels within each instrument are compared against each other using six paired t-tests to affirm that students are interpreting them to be distinct levels of moral reasoning. Specifically, pre-conventional responses were compared against conventional responses, conventional responses were compared against post-conventional responses, and pre-conventional responses were compared against post-conventional responses. Statistically significant results to this test ensured that each level of moral reasoning are perceived as different entities.

In addition, three paired t-tests compare the differences between the mean ratings between the two modes of assessment, one test for each level of moral reasoning. For example, pre-conventional ratings withing the EPSRI were compared to the pre-conventional ratings from Contents Under Pressure. However, running three separate analyses to test a single hypothesis increases the risk of type I error (finding a significant result by chance when in reality there should be a nonsignificant result). Every p-value from the test is multiplied by a Bonferroni correction factor to account for this risk (Bland & Altman, 1995). Cohen's d index is calculated for each test to determine the practical significance between the modes of assessment (Landis & Koch, 1977; Sullivan & Feinn, 2012). Last, the author performed a post hoc statistical power analysis to affirm the sample was large enough to find statistical significance where it may exist (Osborne, 2012c).

Research Question 4.2

The second research sub-question asked, "*how does participation in a digital immersive process safety environment influence the moral reasoning of senior chemical engineering students?*" The author answers this research sub-question by comparing student changes in moral reasoning between the 2018 cohort with the standard curriculum and 2019 cohort with the intervention curriculum. Moral reasoning is evaluated by determining students' N2 scores from the EPSRI. Changes are measured from the beginning of the course (pre-curriculum and pre-intervention) to the end of the course (post-curriculum and post-intervention). The author uses an analysis of covariance (ANCOVA) because differences in moral reasoning may exist between pre-curriculum and pre-intervention. Unlike analysis of variance (ANOVA), ANCOVA considers covariates, or baseline differences between cohorts, and adjusts the post score to account for these differences (Khammar et al., 2020). Next, the author determined η^2 for practical significance (Sullivan & Feinn, 2012) and performed a post hoc statistical power analysis (Osborne, 2012c).

Research Question 4.3

The third research sub-question asked, "*what differences exist in the moral schema of senior chemical engineering students once exposed to a digital immersive process safety environment in comparison to standard process safety instruction?*" This question differs from the previous as it focuses on consistency between the six ordinal moral schema obtained using the CDIT score opposed to just the scale N2 scores. Looking at the schema provides resolution to students' moral development by considering moral consistency. Moral schema are determined using a combination of N2 scores to show advancement in moral reasoning and CDIT scores to show moral consistency (Rest, Thoma, Narvaez, et al., 1997). The author answers this research sub-question by comparing changes in schema using a Stuart-Maxwell test and by comparing nominal changes in moral consistency using a McNemar test. Effect size and statistical power are not reportable for these types of tests.

Results and Discussion

Research Question 4.1

To assure that students were interpreting the three levels of moral reasoning as differentiable constructs within each construct, the mean ratings for each level were compared using paired t-tests. All six continuous variables were checked for normality using skewness and kurtosis and were found to be within normal range, meeting the assumption of the t-test. Six tests were performed to test this hypothesis: pre-conventional to conventional, pre-conventional to post-conventional, and conventional to post-conventional for both of the modes of assessment. Five of the tests were statistically significant ($p<0.001$) and practically significant using Cohen's d for effect size, ranging from a small to large effect (Sullivan & Feinn, 2012). The sixth comparison, between the conventional ($M=4.15$, $SD=0.32$) and post-conventional ($M=4.25$, $SD=0.39$) reflection prompts in Contents Under Pressure, was statistically significant even with the Bonferroni adjustment factor ($\alpha = 6$), $t_{138}=-2.8$, $p=0.035$; however, this test found small practical significance using Cohen's d (Sullivan & Feinn, 2012), $d=0.24$ with a 95% CI [0.07, 0.41]. While this test was adequately powered ($1-\beta =0.91$) and found statistical significance, the Cohen's d confidence interval suggests that replications may find trivial practical differences between students' ratings of conventional and post-conventional reflection prompts. This suggests that students may either find the considerations in these reflections to be similar in construct or in value to their decision making. The remainder of the results suggest students do interpret the levels of moral reasoning to be different, which encourages comparing the levels between the two modes of assessment.

Three paired t-tests were conducted to compare how students valued each of the three levels of moral reasoning built into the second dilemma of the EPSRI and the reflection prompts of Contents Under Pressure. A summary of these differences is shown as a bar chart in Figure 4.4. When comparing the rated importance of pre-conventional considerations (M=3.69, SD=0.70) and reflection prompts (M=3.98, SD=0.67), there was a statistically significant difference, $t_{138}=-4.85$, $p<0.001$. This test found small practical significance using Cohen's d for effect size (Sullivan & Feinn, 2012), d = 0.41 with a 95% CI [0.24, 0.58]. This result suggests that students rated the importance of pre-conventional items slightly higher in Contents Under Pressure and that replications would likely find similar results, ranging from slightly higher to moderately higher. When comparing the rated importance of conventional considerations (M=3.92, SD=0.57) and reflection prompts (M=4.15, SD=0.32), there was a statistically significant difference, $t_{138}=-4.49$, $p<0.001$. This test found small practical significance using Cohen's d for effect size (Sullivan & Feinn, 2012), d=0.38 with a 95% CI [0.21, 0.53]. This result suggests that students rated the importance of conventional items slightly higher in Contents Under Pressure and that replications would likely find similar results, ranging from slightly higher to moderately higher. When comparing the rated importance of post-conventional considerations (M=4.65, SD=0.47) and reflection prompts (M=4.24, SD=0.39), there was a statistically significant difference, $t_{138}=8.38$, $p<0.001$. This test found medium practical significance using Cohen's d for effect size (Sullivan & Feinn, 2012), d=0.71 with a 95% CI [0.52, 0.90]. This result suggests that students rated the importance of post-conventional items moderately higher in the EPSRI than Contents Under Pressure. The confidence interval suggests that replications may find a large effect between the two modes of assessment. Each of these three tests were adequately powered (1-β =1), meaning the sample size was large enough to find statistical significance where it exists.

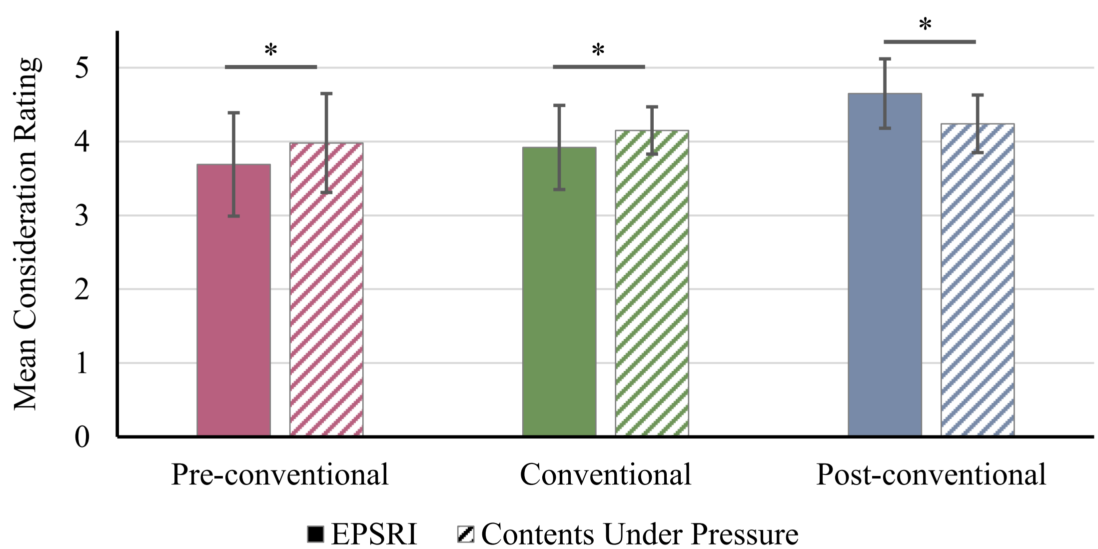


Figure 4.4. Students' mean ratings of considerations between the EPSRI and Contents Under Pressure.

When playing Contents Under Pressure, students' results demonstrated that pre-conventional and conventional considerations had greater value on their decisions than post-conventional considerations. Moreover, the reflections from Contents Under Pressure may better align with the students' expected moral reasoning than their predictions in the EPSRI. Students studying at the collegiate level generally do not reason at the post-conventional level; broad studies performed by Rest et al. (1997) found that individuals in this age and educational range typically operate at the conventional level (concern for those in immediate surroundings), and post-conventional reasoning (concern for society beyond one's circle) is obtained only by those who study moral philosophy. As such, senior chemical engineering students, with limited real-world or industry experience are not likely to reason at the post-conventional level. The students' reflections in Contents Under Pressure may better align with the expected level of reasoning compared to those obtained through rating considerations in the EPSRI.

While both of these modes of assessment focus on process safety decision making, immersion in Contents Under Pressure likely contributes to the differences in students' expressed moral reasoning. In general, the context of a decision may influence one's reasoning to become more predictive of how they should respond (Bazerman & Tenbrunsel, 2011). Instead of reflecting on how one *would* respond, they might instead reflect on how they think they *should* respond. However, immersion may lead to the fading of barriers between the game and its players (Przybylski et al., 2010). Barriers may include distractions caused by controls, external stimuli, or context of play. When making decisions in Contents Under Pressure, immersion may limit the influence of the context (a process safety course) so that responses to dilemmas in the game become more authentic. Specifically, immersion in Contents Under Pressure may influence students to reason authentically at the pre-conventional or conventional level instead of predicting how they think they should reason at the post-conventional level. Researchers have found similar results from qualitative studies among high school students, suggesting games can provide experiences that mimic the real world (Nicaise et al., 2000).

Outside of Contents Under Pressure, it is possible that students respond to ethics instruments in a way that is inauthentic to the level of moral reasoning they may actually use when engaged with a real-world ethical dilemma. The observed difference in moral reasoning between the EPSRI and Contents Under Pressure is not inherently "good" or "bad". As described in Chapter 2, this study does not promote molding engineers into a specific level of moral reasoning. Instead, this work identifies occurrences of behavioral forecasting or predictive mindsets, which is important because when these predictions or forecasts are incorrect, unethical behavior is more likely to occur.

The effect size confidence intervals suggest replications of this experiment are also likely to produce evidence of behavioral forecasting among senior chemical engineering students. As such, process safety instructors should be aware of the possibility that their students may be incorrectly predicting their behaviors. From a pedagogical perspective, the author recommends debriefing students' experiences in Contents Under Pressure with meaningful discussion. In addition, the findings from this study evokes new questions about the impact of Contents Under Pressure and its potential impacts on behavioral forecasting after the immersive experience. These inquiries drive the second research sub-question of this study.

Research Question 4.2

The second research sub-question examines the explicit impact of Contents Under Pressure on moral reasoning by asking, "*how does participation in a digital immersive* *process safety environment influence the moral reasoning of senior chemical engineering students?*" This question is answered through the results of a one-way ANCOVA; the results of which are shown in Table 4.3.

Table 4.3.
ANCOVA Score Adjustment and Results

		Fall 2018 - Standard Curriculum and EPSRI	Fall 2019 - Intervention Curriculum and EPSRI
Baseline N2	M	50.2	62.1
	SD	13.7	16.4
Unadjusted post-course N2	M	51.8	63.0
	SD	15.9	16.6
Adjusted post-course N2	M	56.1	60.4

Both recorded moral reasoning variables were found to be in range of normal distribution using skewness and kurtosis. Levene's test for equality of variances passed and found the two cohorts to be comparable, $F(1,288)=3.47$, $p=0.064$. There was a statistically significant difference in moral reasoning between the two cohorts at the end of the course accounting for baseline moral reasoning scores before the course, $F(1,287)=5.97$, $p=0.015$. This test found a small effect size, $\eta^2=0.02$. This suggests that, while there is a statistical difference in moral reasoning between the comparison and experimental cohorts, the difference is of little practical significance. A post-hoc power analysis found this test was underpowered ($1-\beta = 0.68$). Although this test did obtain statistical significance, replications of this work should aim for a larger sample size to avoid the risk of rejecting a false null hypothesis.

When looking at the baseline N2 scores, the intervention cohort had a notably larger average than the comparison cohort, which affirmed the need for an analysis that accounts for covariates. While it is not known why the intervention cohort had a larger baseline N2 score, the difference was likely driven by biases outside of the scope of the study, such as ethics interventions outside of the classroom or instructors adapting their explanation of the EPSRI between implementations.

In Table 4.3, both the comparison and intervention cohorts show an increase in their unadjusted N2 scores; this result may be occurring in-part due to the cohorts responding to the EPSRI a second time as they become familiar with the instrument. Mayhew (2015) claims that responses to ethics instruments will likely increase in score because respondents have responded more than once. As such, the mean N2 scores may be slightly inflated because of retesting.

Playing through Contents Under Pressure has limited practical significance on senior chemical engineering students' N2 scores. The N2 describes an individual's moral reasoning by the prevalence of their post-conventional reasoning in the absence of pre-conventional reasoning (Rest, Thoma, & Edwards, 1997; Rest, Thoma, Narvaez, et al., 1997). The N2 score is a standard metric in discussing an individual's moral reasoning (Rest, Thoma, & Edwards, 1997; Rest, Thoma, Narvaez, et al., 1997), yet this value speaks nothing about the extent of an individual's conventional reasoning because it is not considered when calculating N2 scores. To gain deeper understanding on the intervention, the author also evaluates these students' moral schema through the third research sub-question of this study. While not the standard in assessing moral reasoning, this may add finer resolution to these students' moral development by also considering their conventional reasoning and consistency to the three levels.

Research Question 4.3

The third research sub-question of this study asked, "*what differences exist in the moral schema of senior chemical engineering students once exposed to a digital immersive process safety environment in comparison to standard process safety instruction?*" The schema in this research sub-question refers to Rest et al.'s (1999) neo-Kohlbergian approach to moral development that considers moral consistency. To answer this research sub-question, the author first performed a Stuart-Maxwell test; the crosstabulation from this test for the comparison and intervention cohorts are shown in Table 4.4 and

Table *4.5*. To visually compare changes to these cohorts, representations by schema are normalized as percentages in Figure 4.5. The crosstabulation of the comparison cohort shows that most students were consolidated in their post-conventional reasoning (Schema 6) and remained here (n=31). This is followed by students who were transitional in their post-conventional reasoning (Schema 5) and either remained in that schema (n=21) or moved to be consolidated in that level (n=17). The Stuart-Maxwell test found that these changes were not statistically significant, p = 0.752. The crosstabulation of the intervention cohort shows that the majority of students remained in their post-conventional schema as either consolidated (n=86) or transitional (n=32). The next largest group of students reasoned post-conventionally going from consolidated to transitional (n=24). This change, among others, promoted the Stuart-Maxwell test for this cohort to yield statistical significance, p = 0.013. While Figure 4.5 shows that both cohorts experienced similar percentages of change per schema, proportionally more students shifted their schema in the intervention cohort than the comparison cohort.

Table 4.4.
Crosstabulation and Summary of Schema Changes for the Comparison Cohort (n=109)

Comparison		Post-course					
		1, C	2, T	3, T	4, C	5, T	6, C
Pre-course	1, C	0	0	0	0	0	0
	2, T	0	1	0	0	2	0
	3, T	0	1	3	0	4	4

Comparison		Post-course 1, C	2, T	3, T	4, C	5, T	6, C
	4, C	0	0	0	0	1	1
	5, T	0	3	7	1	21	17
	6, C	0	0	2	1	9	31
Change to schema		±0	+2	±0	±0	-12	+10

Note. C denotes consolidated moral reasoning; T denotes transitional moral reasoning.

Table 4.5.
Crosstabulation and Summary of Schema Changes for the Intervention Cohort (n=181)

Intervention		Post-course					
		1, C	2, T	3, T	4, C	5, T	6, C
Pre-course	1, C	0	0	0	0	0	0
	2, T	1	3	2	0	1	1
	3, T	0	1	0	0	4	0
	4, C	0	0	0	0	1	2
	5, T	0	2	9	0	32	9
	6, C	0	1	2	0	24	86
Change to schema		+1	-1	+8	-3	+10	-15

Note. C denotes consolidated moral reasoning; T denotes transitional moral reasoning.

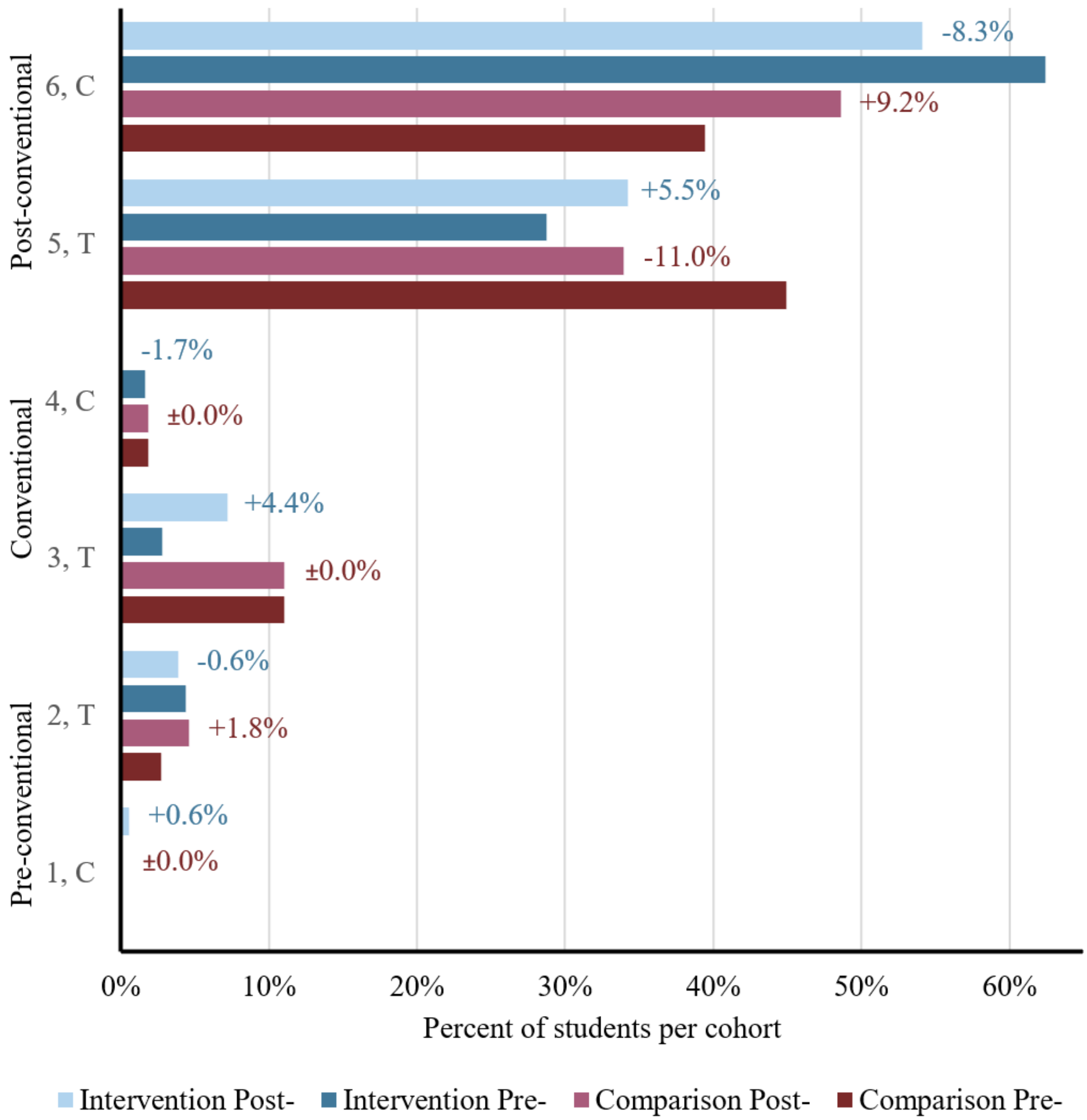


Figure 4.5. Bar chart of schema representation and changes.

The results from the crosstabulations showed how a majority of the students whose level of moral reasoning changed did so between transitional and consolidated schema at the post-conventional level. To supplement the Stuart-Maxwell analysis, the author performed a McNemar test to see how students move between transitional and consolidated schema. The crosstabulation for the comparison cohort is shown in Table

4.6, and a McNemar test found these changes in moral consistency to not be statistically significant, p = 0.123. The crosstabulation for the intervention cohort is shown in Table 4.7, and a McNemar test found these changes to be statistically significant, p = 0.010. While both of these cohorts experienced similar changes in moral schema, these results show that the intervention cohort became more morally transitional.

Table 4.6.
Crosstabulation of Moral Consistency Changes for the Comparison Cohort

		Post-course	
		Consolidated	Transitional
Pre-course	Consolidated	33	12
	Transitional	22	42
Changes to moral consistency		+10	-10

Table 4.7.
Crosstabulation of Moral Consistency Changes for the Intervention Cohort

		Post-course	
		Consolidated	Transitional
Pre-course	Consolidated	88	28
	Transitional	11	54
Changes to moral consistency		-17	17

As a result of playing Contents Under Pressure, students expressed moral reasoning across more levels of development when responding to the EPSRI. Between the two cohorts, the intervention cohort statistically significantly became more transitional in their reasoning. Rest et al. (1999), who developed the neo-Kohlbergian extension to moral development, describe transitional consistency as reasoning with a preference towards one level yet reasoning with other levels evident. For example, students who were developed at the fifth schema were predominantly reasoning post-conventionally, but there was also evidence of conventional level reasoning. While Kohlberg's framework suggests development at higher levels allows lower levels to still be accessed, consolidated reasoners do not necessarily access them. The students from the intervention pulled from multiple levels of reasoning when responding to these ethical dilemmas, unlike the students who were not exposed to Contents Under Pressure. In addition, because the first research sub-question of this chapter suggests these responses may be wrongly predictive of future behaviors and because the comparison cohort did not significantly change over the semester, it is possible that both the pre- and post-course responses from the comparison cohort are inaccurate predictions. To reiterate a common theme, the goal of this work is not to push students towards a specific level or schema of moral reasoning. However, these results suggest that Contents Under Pressure influenced students to apply more perspectives to when responding to ethical dilemmas.

Individuals who morally reason transitionally use a blend of levels of moral reasoning, which is more complex than a single level. As shown with the model of moral reasoning (Figure 2.2), changes in schema can be gradual; thus, reasoning transitionally may provide evidence of such gradual development (Derryberry & Thoma, 2005; Thoma & Rest, 1999). Moreover, literature suggests individuals in a senior level engineering course are unlikely to reason at the post-conventional level, especially with consolidated consistency (Rest, Thoma, Narvaez, et al., 1997). When either of the cohorts' responses to the EPSRI found evidence of students reasoning at Schema 6, it is safe to infer these students may have been responding with a predictive mindset (Bazerman & Tenbrunsel, 2011). However, when the intervention cohort became more transitional, especially with the 5.5% increase over Schema 5 (Figure 4.5), these students may be reasoning less predictively and more authentically because of the gained experience from the game (Barab & Duffy, 2012).

Results from Research Sub-question 4.2 found statistical significance but small practical differences between the comparison and intervention cohorts. This result suggests there is some difference between the cohorts, which can be attributed to the large sample size, but there is not a meaningful takeaway. However, when investigating changes by schema and moral consistency, the intervention cohort displayed statistically significant changes (there are no effect size metrics appropriate for Stuart-Maxwell tests). This difference becomes evident when observing the crosstabulations in Table 4.6 and Table 4.7; very few students moved in or out of predominantly reasoning at the pre-conventional level. The majority of changes across the semester occurred within the conventional and post-conventional levels. As such, the N2 score commonly accepted as a standard in assessing moral development from ethics interventions may miss out on critical development when used alone.

Contents Under Pressure, in its present format, may be limited in its impact on students' moral development because it currently only spans 15-days of gameplay. When differentiating the levels of moral development, Rest et al. (1997) describe advancement by age brackets and degrees of education (secondary, undergraduate, and graduate school). They go on to describe that individuals shift in moral development gradually and over long periods of time. While ethics instruments may be sensitive enough to measure differences in moral development as a result of an intervention (Rest, Thoma, Narvaez, et al., 1997), a brief intervention likely will not merit extensive advancements or changes to one's moral reasoning. Contents Under Pressure may influence how senior chemical engineering students approach ethical decisions in process safety contexts, but because it is only a 15-day intervention, it may be limited to the extent it can change how these students reason. It is possible a longer immersion in the game through an extended narrative may have greater influence on students' development, but further testing with a longer duration intervention would be required. Note, Contents Under Pressure is intended to be used as a pedagogical tool. While this study compared responses between Contents Under Pressure and the EPSRI, Contents Under Pressure was not designed to be used like a survey instrument.

Generalizability and Limitations

The three research sub-questions of this study are only answered using data from senior chemical engineering students from three US ABET accredited chemical engineering programs. While statistical significance and power may provide confidence and reproducibility in the sampled group, the takeaways may lack broad generalizability. The author chose to work with senior chemical engineering students because they are the population that is typically targeted for process safety education (Mkpat et al., 2018), yet these findings have limited implications for programs that teach process safety earlier or throughout the chemical engineering curriculum. Focusing on university students also means that Contents Under Pressure has limited implications towards industry practitioners who may be seeking additional process safety education. This study only samples from US-based institutions, so there may be limitations to how it would impact students from other countries. For example, Clancy (2017) compared moral reasoning between US and Chinese institutions and found that cultural differences are significant. Thus, further work is required to generalize these findings to other regions. Moreover, these data do not consider the impact of gender or ethnicity. While there is no theoretical basis for considering these variables (Osborne, 2012a), it is possible that nuances in students' moral reasoning may exist in relationship with aspects of an individual's personal identity. Lastly, this work is focused on chemical engineers. While Contents Under Pressure has been designed to not require in-depth knowledge about chemical processes and plants, it is unclear how Contents Under Pressure may impact the moral reasoning of other engineering disciplines or non-engineers.

For the first research sub-question comparing students' expressed moral reasoning between the considerations of the two modes of assessment, Contents Under Pressure is compared to only the second dilemma from the EPSRI because of similarities between the narratives. Thus, moral reasoning may have some variance between this dilemma and the remaining dilemmas in the EPSRI. In addition, Contents Under Pressure's narrative varies slightly based on the player's decisions, creating some dynamic pathways. Because of this, some reflection prompts are not seen by all students, which may have had some influence on the average moral reasoning scores. For the second and third research sub-questions, participants from both cohorts responded to the EPSRI twice. Work by Mayhew et al. (2015) found that repeat attempts with ethical testing instruments inherently lead to increased post-conventional scores. It is possible that the findings from this work are biased because they are responding to the EPSRI twice. Finally, student experience with games may have influenced students' moral reasoning within and from Contents Under Pressure. Studies on digital games have acknowledged that prior experience with games and controls influences gameplay (Dörner et al., 2016; Orvis et al., 2010; Przybylski et al., 2010; Whitson et al., 2008). Though this study does not consider students' prior experience with games or mastery of games, it is possible that students with more game experience were more influenced by Contents Under Pressure than those with less game experience.

These research questions used the version of the EPSRI available at the time of the study (EPSRI V_4) described in Instrument Development in Chapter 3. While a single version of the EPSRI was used across these research questions, an updated version (through the final round of EFA and CFA, discussed in the previous chapter) is now available (EPSRI V_6). While there is sufficient evidence to have confidence in the validity of these results, it is possible replications of this work may find some variance due to using an updated version of the EPSRI.

Finally, it should be noted that students using ad blockers in their internet browsers may have been omitted from data used in Research Sub-question 4.1 because Google Analytics was unable to record their decisions in Contents Under Pressure. However, it is unlikely this may have impacted the representation of the sampling given the large sample size (n=148).

Conclusions

The objective of this study is to assess the impact of a digital process safety game, Contents Under Pressure, on the moral reasoning of senior chemical engineering students. When approaching ethical dilemmas, ethical fading or inaccurate behavioral forecasting holds the potential to lead to unethical behavior. The overarching research question for this study asked, "*how does a digital, process safety game influence the ethical decision making of chemical engineering students?*" Building awareness surrounding poor predictions and ethical fading may help mitigate one's ethical blind spots (Bairaktarova & Woodcock, 2017; Clancy, 2020a; Osberg & Shrauger, 1986). The game provides senior chemical engineering students with an opportunity to test their ethical decision making in authentic situations (Barab & Duffy, 2012), which may be assessed through their moral reasoning. The game is assessed through both a retrospective analysis and a cross-sectional analysis using quantitative methods.

The cross-sectional analysis compares students' perceived value of each level of moral reasoning by their ratings of dilemma considerations in the EPSRI and rated importance of reflection prompts in Contents Under Pressure. Only a single dilemma from the EPSRI is used to affirm narrative similarities between the two modes of assessment. The retrospective analysis compares two student cohorts' N2 scores, predominant level of reasoning, and CDIT score. Differences between the comparison and intervention cohorts may be attributed to Contents Under Pressure acting as an ethics intervention in the students' process safety curricula. The N2 score represents the prevalence of post-conventional moral reasoning in the absence of pre-conventional reasoning. The CDIT score is a benchmark of moral consistency, whether it is consolidated or transitional to a level of moral reasoning. Last, the predominant form of moral reasoning is considered to be the most exerted level of moral reasoning (despite usage of other levels of reasoning).

This study found that Contents Under Pressure recorded moral reasoning at a level more consistent with expectations in literature than the EPSRI (Rest, Thoma, Narvaez, et al., 1997). **This result suggests that, as an intervention, Contents Under Pressure can reduce the influences of behavioral forecasting and predictive mindsets**. It is possible that the immersive attributes of the game led students to respond in an authentic manner, whereas responses to the EPSRI may be biased by social desirability (Edwards, 1953) and behavioral forecasting (Bazerman & Tenbrunsel, 2011). When evaluating the impact of Contents Under Pressure on students' moral reasoning, the EPSRI did not observe differences in students' N2 scores. However, when looking at their moral schema, this study found evidence that playing Contents Under Pressure alongside process safety curricula influences how students would respond to the EPSRI with transitional moral consistency. Transitional moral consistency means students are applying more perspectives to responding to ethical dilemmas and are becoming prepared to make advancements to their moral development (Derryberry & Thoma, 2005; Rest, Thoma, & Edwards, 1997). In answering the overarching research question of this study, the process safety game, **Contents Under Pressure, can yield more authentic decision making, likely through its immersive context**, yet significant advancements to moral reasoning are not recorded in the game's current format. To directly answer the overarching research question, a digital, process safety game influenced the ethical decision making of chemical engineering students to be authentic by mitigating behavioral forecasting.

This study supports the use of Contents Under Pressure in process safety curriculum because it challenges students' behavioral forecasting. When process engineers inaccurately predict their behaviors and decisions, it is possible that they make decisions that could have negative implications in a plant based environment. Contents Under Pressure provides students with an opportunity to practice their decision making and recognize what influences their decisions. In addition, changes in moral development as a result of the game experience may lead to discussions on how to approach ethical dilemmas with an awareness of one's own blind spots.

Chapter 5

Engineers' Criteria Founded Decision Making

Overview

The objective of this third study is to understand how safety is considered among an array of process safety criteria. Specifically, the criteria considered in this study include those in the conceptual framework described in Chapter 2 (Akinleye et al., 2019; Dönmez & Uslu, 2020; Encinosa & Bernard, 2005; Hendrickson & Au, 2008): (1) budget, (2) personal relationships, (3) plant productivity, (4) safety, and (5) time. As discussed in the literature review, these criteria demonstrate importance in engineers' approach to decision making, yet to what degree is unclear. This study is guided by the overarching research question: *what role does competing criteria hold in the process safety decisions of engineering students?*

This study considered senior chemical engineering students' claimed criteria priorities, explanations for these priorities, and actual priorities as shown through decisions made within a digital process safety decision making game, Contents Under Pressure. First, students express priorities by ranking the aforementioned criteria at the beginning and end of their course through a survey. Students also reflect on how they came to these rankings through a written response. Next, during the course, students play Contents Under Pressure; decisions made in the game were evaluated for their impact on in-game metrics similar to process safety criteria. Results broadly showed that safety is prioritized across the three data sources, yet the results suggest students may be inexperienced in making decisions with competing process safety criteria.

Introduction

Contents Under Pressure

This study leverages a digital educational game designed for chemical engineering students called Contents Under Pressure. This game allows players to take on the role of a chemical plant manager who must make binary decisions to balance an array of game metrics: time, safety, reputation, and plant output. The game has a 15-day narrative of events in which the player will interact with employees and supervisors. Further description on Contents Under Pressure is provided in Chapter 4.

Paragaming

Paragaming is a gameplay strategy that focuses on achievement hunting or playing solely to obtain high scores instead of engaging game narrative or assigned objectives (Carter et al., 2012). For example, in Rockstar's game Grand Theft Auto (GTA) V, players role-play as members of an organized crime group, yet the game provides achievements for non-narrative focused tasks, such as mastering hobbies like tennis or darts (Figure 5.1). Intentionally seeking out achievements outside of the game narrative may suggest that players are not immersed in the story (Carter et al., 2012). In the context of Contents Under Pressure, when a player paragames, their decisions focus on optimizing outcomes, which may lead to inauthentic decisions from the diminished immersion. Earlier studies in motivation within Contents Under Pressure identified paragaming as a theme of students' desired game outcomes (Stransky et al., 2021).

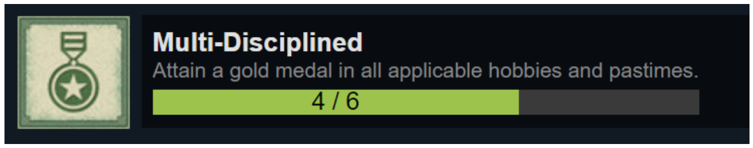


Figure 5.1. Achievement offered in GTA V for mastering hobbies (Rockstar Games, 2013).

Study 3 Research Questions

This study aims to answer one overarching research question, *what role does competing criteria hold in the process safety decisions of engineering students?* To do so, the author answers three research sub-questions as building blocks to the overarching question:

5.1 How does senior chemical engineering students' prioritization of process safety criteria change after exposure to a digital process safety game?

5.2 How do senior chemical engineering students explain their rankings of process safety criteria?

5.3 How do senior chemical engineering students' decisions within a digital process safety game indicate their prioritization of process safety criteria?

Overview of Methods

Data were collected from senior chemical engineering students from four chemical engineering programs using Contents Under Pressure and pre- and post-game reflection surveys. Criteria rankings were compared using an ANOVA, and explanations were evaluated through thematic analysis. Students' gameplay was evaluated using time-series analysis.

Overview of Results

This study found that students prioritized safety through both their reflections and gameplay decisions. Out of the five criteria, students initially ranked the personal relationships criteria as their lowest priority, but after playing Contents Under Pressure, where reputation was shown to be of importance, personal relationships statistically significantly increased in ranked importance. When justifying their rankings, students primarily recalled on prior experiences, aligned their decisions with business responsibilities, or linked criteria together to describe importance. While safety may have been ranked as most important overall, these results do supply evidence that students may yet be inexperienced in ranking competing criteria.

Methods

Study 3 Design

A total of 287 senior chemical engineering students participated in the study from four chemical engineering programs, which were chosen as part of the author's professional network. As Study 3 focuses on connecting the findings among the three research sub-questions, the author determined that single, consistent sample was necessary across the data sources for each sub-question. If a response was missing from any data source, the data were removed. While some random missing data may be interpolated (Osborne, 2012b), incomplete responses were removed because qualitative responses and in-game decisions cannot be statistically interpolated. A shared sample across the three research sub-questions enables more robust claims by contextualizing answers with multiple data sources. Responses were also removed when Contents Under Pressure data showed evidence of paragaming. In this study, evidence of paragaming may resemble a single student who simultaneously plays the game twice, searching for the best outcome from each decision. After filtering, the final sample used for data analysis consisted of 82 senior chemical engineering students.

Research Question 5.1

The first research sub-question of this study asked, "*how does senior chemical engineering students' prioritization of process safety criteria change after exposure to a digital process safety decision making game?*" This research question was answered by analyzing how students ranked the five process safety criteria from a pre- *and* post-game reflection survey.

- Rank / order the criteria listed (budget, personal relationships, plant productivity, safety, and time) based on how important they are to your decisions within a chemical plant environment. The most important criteria should be at the top of the list and the least important should be at the bottom.

Differences among how criteria were ranked within each survey were compared using a repeated measure ANOVA. An ANOVA post-hoc test determined specific sources of statistical significance. Cohen's d index provided insight on the practical significance between each pair of criteria (Sullivan & Feinn, 2012). Last, a post-hoc power analysis was performed to affirm the sample was large enough to find statistical significance where it may exist (Osborne, 2012c).

Research Question 5.2

The second research sub-question sought to gain understanding in how the reasoning of criteria priorities develop through experiences. Specifically, this question asked, "*how do senior chemical engineering students explain their rankings of process safety criteria?"* This research question was answered by performing a thematic analysis on students' written reflective explanations for their criteria rankings during the pre- and post-game surveys, "*Provide an explanation for the rank / order you have selected above and how this prioritization will impact your approach to decisions within a chemical plant environment.*"

Two researchers separately read all of the students' reflections from both pre- and post-gameplay at once and took notes on their observations, consistent with the methods described by Saldaña (2016). Upon completion, they developed codes and code descriptions that captured the data until theoretical saturation was reached (Charmaz, 2006). These two researchers compared their initial findings, debated which codes best represented the data, and consolidated their findings into a few latent themes in a code book.

Next, researchers separately thematically coded on a subset of 25 reflections to confirm coding agreement between the two researchers, where multiple themes could be applied to a single reflection. Inter-rater coding agreement of the reflections was evaluated using Cohen's inter-rater reliability kappa statistic, κ (Landis & Koch, 1977; Miles et al., 2014). Reliability below the acceptable threshold ($\kappa > 0.70$) was obtained, so the researchers completed coding the whole sample and discussed discrepancies. Where discrepancies reoccurred in certain themes, the codebook was amended to improve code clarity. Throughout this process, the researchers kept a living audit trail to maintain record of their work and affirm process reliability (Lietz et al., 2006; Walther et al., 2017).

Quality Concerns

As this work considers qualitative data, the author leverages the quality framework developed by Walther et al. (2013, 2017) to ensure validity and reliability in the results. A summary of the quality attributes in this framework and how they were addressed in this work is shown in Table 5.1.

Table 5.1.
Quality Framework Applied to this Study (Walther et al., 2013, 2017)

Quality Attribute	Description	How its obtained
Theoretical validation	Ensures that context under investigation aligns with the theoretical or conceptual framework used in the evaluation	By developing a conceptual framework around relevant sources of literature (Walther et al., 2013) and considering literature in the interpretation of the results.
Procedural validation	Ensures fit between aspects of the research design and the theory involved.	Through two primary methods. (1) The author' s codebook development and analytic decisions with justifications were captured through an audit trail, and (2) two researchers coded all student reflection data and reconciled differences.
Communicative validation	Ensures that findings are communicated in a way that is in accordance with the knowledge and meaning of the research community.	By examining the wording and meaning of all codes with a researcher outside of the author.

Quality Attribute	Description	How its obtained
Pragmatic validation	Ensures that the theoretical and conceptual framework are compatible with observations in the context under investigation.	By examining the findings for implications towards educational and industrial practice around process safety decision making.
Process reliability	Ensures that the analysis is independent from random and unpredictable influences.	By distributing surveys in consistent manners and recording the analysis through an audit trail.

Research Question 5.3

The third research sub-question of this study asked, "*how do senior chemical engineering students' decisions within a digital process safety game indicate their prioritization of process safety criteria?*" The author answers this question by looking at the actual decisions students made while playing *Contents Under Pressure*. In the game, students must make decisions while accounting for game metrics time, safety, reputation, and plant output. While these game metrics differ from the five criteria described in the conceptual framework, there are indeed similarities between the two. By understanding how students balance, prioritize, and discount the metrics through their actual actions in an authentic context, it provides some insight on how they might handle the need to balance process safety criteria in the real world. This component of the study leverages the fundamentals of Time Series Analysis (TSA). TSA plots provide a visual of changes in data over time, and has typically been applied to psychology studies to find the impact of an intervention (Velicer et al., 2012) as well as to economic trends to make predictions. However, TSA has increasingly been used in game studies searching for player types and pathway trajectories (Reilly & Dede, 2019; Sawyer et al., 2018; Zuparic et al., 2021). In the case of *Contents Under Pressure*, the author plotted students' metrics scores over the course of the game narrative to observe the trends, abnormalities, and uniqueness in the game metric scores. TSA with mean metric scores and standard deviation error bands shows the mean relationship among the safety, reputation, and plant output as a form of answering the third research sub-question. Specifically, analyzing time series data provides insight on how chemical engineering students approach making process safety decisions which have competing criteria.

Results and Discussion

Research Question 5.1

To answer the first research sub-question, “how does senior chemical engineering students’ prioritization of decision-making criteria change after exposure to a digital process safety decision making game?”, the author performed a repeated measure ANOVA. When asking students to rank the importance of process safety criteria at the beginning and end of their course, most of the criteria underwent statistically significant changes in importance. A summary of these results is shown in Table 5.2. Two of the tests in the table were adequately powered: the budget and productivity criteria. The ranking of the budget criteria moved down in importance with statistical significance and a large effect size. The ranking of the productivity criteria remained the same as the test did not obtain statistical or practical significance. Three of the tests were underpowered, indicating the potential need for replications of this work with a larger sample. The ranking of personal relationships moved up in importance as the test found both statistical and large practical significance. The ranking of the safety criteria remained the same. While this test was under powered when not finding statistical significance, the effect size suggests that there was no practical change to this ranking. Last, the ranking of the time criteria did move down in importance slightly, approaching statistical significance. This suggests that it was possible to obtain statistical significance if a larger sample size were obtained. However, the effect size showed the change is of small practical significance; while statistical significance could be obtained with a larger sample, the difference would likely not add more value in answering the research sub-question.

Table 5.2 .
Summary of Results from ANOVA (n=83)

	Budget		Productivity		Personal Relationships		Safety		Time	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Mean ranking	2.98	3.77	2.71	2.48	4.25	3.27	1.20	1.30	3.86	4.18
Standard deviation	0.91	0.89	1.08	0.96	1.08	1.23	0.54	0.64	1.06	1.11
Z statistic	Z = -5.12		Z = -4.71		Z = -1.79		Z = -1.10		Z = -1.95	
Statistical significance	$p < 0.001$		$p = 0.074$		$p < 0.001$		$p = 0.274$		$p = 0.052$	
Effect size (Sullivan & Feinn, 2012)	d = -.67 Medium effect		d =.19 Small effect		d = 0.61 Medium effect		d = -0.13 No effect		d = -0.21 Small effect	
Statistical power (Osborne, 2012c)	$(1-\beta) = 0.99$ Adequately powered		$(1-\beta) = 0.99$ Adequately powered		$(1-\beta) = 0.37$ Under powered		$(1-\beta) = 0.37$ Under powered		$(1-\beta) = 0.60$ Under powered	

The plant productivity and time criteria generally held the same ranked importance across game play. The lack of change in ranking may suggest that either students' perception of these criteria did not change or that while their perception of the criteria changed, with respect to their dynamic perception of other criteria, they remained ranked the same. A major finding from study 2 in Chapter 4 found that 15-days of immersion in Contents Under Pressure may have begun preparing students to make changes in their moral development, which indicates that the incremental and small changes in criteria ranking may be amplified given a longer immersion duration. Claims explaining the budget criteria decreasing in ranked importance cannot be made yet, despite the large effect size, until paired with qualitative data. However, this decrease in how students ranked its importance must be distinguished from its actual decrease in importance; this lower ranking does not necessarily mean that the criteria is "unimportant" to students' process safety decisions.

At the end of the semester, students significantly increased the ranked importance of the personal relationship criteria, with a medium effect size. While this change cannot yet be explained without sufficient qualitative insight, literature around decision making and relationships provides potential explanations. It is possible that, from the game immersion, students were able to obtain more satisfaction when work objectives were balanced with personal relationships. Friedman and Greenhaus (2011) describe that a healthy work-life balance increases the capacity for satisfaction and happiness; thus, when students prioritized the personal relationships criteria more, they were able to obtain more satisfaction from their work. In contrast, an increase in prioritizing this criteria may not have been driven by seeking satisfaction but by avoiding dissatisfaction. Contents Under Pressure often juxtaposes these five criteria against each other. Early in the game, the player's supervisor, Wanda, asks the player to stay late after work to finish a report. Players who agree with the goal of appearing dedicated to the job and building reputation with their boss receive game feedback, "*Wanda is happy you're putting in extra work, but you miss your family's surprise dinner to celebrate your first day*." Thus, relationships built with in-game characters may have influenced students to increase the ranked importance of the personal relationship criteria to avoid negative outcomes.

This situation is not unique to Contents Under Pressure as process safety case studies have documented how personal relationships influence decision making. An explosion at a Chevron refinery was exacerbated when a team of engineers identified the plant needed to be shut down to isolate a leaking pipe, yet none made the call for fear of "flack" from their colleagues (CSB, 2015). Familial relationships led to deteriorated safety culture at an Aghorn Operating Inc. waterflood pump station in Odessa, Texas. This poor culture led an employee's partner to fatally enter a restricted area during a chemical leak (CSB, 2021; Geddes, 2019). It is likely that senior chemical engineering students do not have sufficient experience in managing work and familial relationships when they compete with work objectives. The increase in ranked importance of the personal relationship criteria as they gain relevant experience supports this claim. Thus, it becomes increasingly important for students to build an awareness of how relationships may influence their decision making, such as through the Contents Under Pressure experience.

Last, the ranked importance of the safety criteria experienced a negligible change according to the effect size (Cohen, 1988; Sullivan & Feinn, 2012). This shows that despite other criteria moving around in importance, at the end of the course, their perceptions of the importance of safety generally remained the same. Specifically, out of the 82 of students who ranked safety as number one during the pre-course survey, 57 of them (69%) continued to rank safety number one in the second survey. However, it is possible that some of these responses are biased because of the context. In Chapter 4, we discussed how behavioral forecasting (Bazerman & Tenbrunsel, 2011) may have promoted students to respond more post-conventionally than they actually morally developed possibly because they were aware they were in the context of an ethics assessment. **Similarly, students may be ranking the safety criteria highly because they are aware that they are responding to a survey in the context of a process safety or a senior design course or after having recently participated in a process safety game.** This may also be explained through a phenomenon known as social desirability bias (Edwards, 1953). This bias occurs when an individual behaves in a way consistent with how they believe others in the environment would want them to behave with the goal of being likeable (CSB, 2015). This could lead to senior chemical engineering students responding to this survey in a way they believe is consistent with what the instructor may want out of them, thus ranking the safety criteria highly. It is unclear to what extent these biases may be at play or to what extent the students were responding earnestly.

Changes in students' ranked priorities may be limited by the list of five criteria provided to them. Additional criteria or nuances to the criteria may exist, such as differentiating personal relationships from work relationships. Moreover, students may be influenced to value these five criteria because they were provided in the survey potentially implying their importance, whereas students introducing criteria as their own process safety priorities may lead to an alternate outcome.

In answering the first research sub-question, the data only provided evidence that the personal relationships and budget criteria changed due to participation in the game. It is unclear why budget moved down in ranked importance. Personal relationships moved up in importance possibly because the students drastically lacked experience making engineering decisions with this criteria before the course and game experience. The remaining three criteria (plant productivity, safety, and time) did not significantly change in importance.

Research Question 5.2

The second research sub-question asked, "*how do senior chemical engineering students explain their rankings of process safety criteria?*" The author answers this research question through thematically coding students' written responses to the prompt, "*Provide an explanation for the rank / order you have selected above and how this prioritization will impact your approach to decisions within a chemical plant environment.*" Upon reviewing the data from both pre- and post-game reflections, the author developed a codebook with eight themes on students' explanations, shown in

Table 5.3. These codes are also described in detail in the following subsections. Multiple themes emerged from single reflection as students provided multiple forms of justification within their reflections.

Table 5.3.
Codebook Used to Thematically Code Student Reflections in Answering Research Sub-question 5.2 (n=83)

Students explain criteria rankings based upon-	Code	Description	Example
-the criteria' s importance	with explicit certainty	Students express the importance of a criteria with explicit certainty. Reasoning for its importance is likely limited in description.	"Safety is of utmost importance, as without safety, there is nothing. Time is…"
	with flexibility or contextualization	Students describe that the importance of a criteria may be conditional on other criteria, that the importance is contextual or flexible, or that the outcome of deprioritizing the criteria is remediable. This does not necessarily include the "least" important criteria in a list.	"...Time and budget can be the most flexible. Unfortunately, sometimes good work can take a little longer."
	not being considered	Students deny that a criteria is of importance or suggest that the criteria might be a hindrance to process operations.	"...Finally, personal relationships should not be considered within a chemical plant environment, as it is a place of work and should be kept a professional setting."

Students explain criteria rankings based upon-	Code	Description	Example
-the connection	to held personal values	Students express a criteria contributes to maintaining their personal values, such as security, loyalty, or status.	"Safety is always my top priority because for one, <u>all plant workers need to feel safe, secure, and confident in their workplace</u>... Personal relationships is fourth because <u>I believe that workplace relationships are very important</u>. Trust and loyalty go a very long way..."
	to other criteria	Students justify ranking a criteria by its connection to another criteria. Students may draw criteria connections through a criteria's relationship or influence on other criteria. Students may also not discern differences between criteria.	"My first two are close as plant productivity should be the goal of every chemical engineer, to maximize plant efficiency... <u>Budget is next because it ties in with plant productivity and the two are heavily related</u>…"

Students explain criteria rankings based upon-	Code	Description	Example
-their goal	to avoid potentially negative outcomes	Students rank criteria with the goal of avoiding negative process outcomes such as injuries or environmental repercussions.	"Safety is the most important factor because an unsafe environment can cause unnecessary risk, injury, death, lawsuits, or other events that should be avoided at all costs…"
	to support business responsibilities	Students rank criteria with the goal of supporting company responsibilities or labor objectives, dealing in facts and	"The plant is there to make money; they have a duty to the shareholders of the company to provide product, so productivity is number one…"
-experience		Student calls on prior experience from their course or gameplay to justify their ranking of a criteria.	"I focused mainly on the safety aspect so, if possible, I would always make sure that stayed in the top two ratings during the game... The budget wasn't a big focus with Contents Under Pressure as it really didn't affect too much during the game, so I thought about that the least."

Student Explain Criteria Rankings based upon the Criteria's Importance with Explicit Certainty. While describing the importance of a criteria, students would describe the criteria to be important with limited explanation. Many times, they would not stray from its importance, and often they would emphasize it was a priority that could not be changed. The lack of explanation suggests that students had a firm determination or resolute perspective on the criteria. The following quotes are examples of how students explained their rankings by establishing the criteria's resolute importance

- "Time is of the upmost concern in any manufacturing facility…"
- "Safety is #1. Getting everyone home safely every day is what matters most."

Student Explain Criteria Rankings based upon the Criteria's Importance with Contextualization. Students also described that the criteria's importance could change depending on specific conditions. Students would create specific situations where their opinion on a criteria may change or define a criteria to be important but willing to trade it for other criteria. The following two quotes provide examples of how this theme occurred in the data.

- "-Next, I see my time as valuable and ensuring that the work I complete will be worthwhile and not wasted. Thus, <u>assuming that the issue is not urgent or catastrophic in nature, I would prioritize it appropriately</u>… <u>Lastly, my personal relationships would only be consorted if absolutely necessary</u> as I prefer to keep my work life and social life separate. <u>If an</u>

issue was catastrophic enough it might affect them, I would prioritize letting them know."

- "While I did attempt to keep the productivity high, I found that it was normally a much lower priority when it came to choosing between it and safety and relationships…"

Student Explain Criteria Rankings based upon the Criteria's Importance with Not Being Considered. Some students would deny the importance of certain criteria when prompted to explain their rankings of criteria in the conceptual framework. Some would go as far as to explain that the criteria held no place in their decision making model whatsoever. While the conceptual framework found evidence in literature that these criteria may be of general value to engineers' decision making, it is not impossible for these criteria to be excluded among some decision makers. The following quotes are examples of students who explained ranking personal relationships last by denying its importance in the workplace.

- "Last, personal relationships should be the least important in decision making; while I consider them important, favoritism and interpersonal relationships should be ignored as much as possible in a critical situation in favor of maintaining clear focus of what needs to be done."
- "-I put relationships and time at the bottom because from a work perspective those are not priority even though they are priorities in life."

Student Explain Rankings by Connecting to Held Personal Values. While describing the importance of some criteria, students would pull on personal values of theirs. Lichtenstein et al. (2017) describe personal values as a series of beliefs or concepts that guide one's actions or considerations while making a decision. These personal values may include justice, honesty, reputation, kindness, etc. Students considering their personal values in making decisions is consistent with literature on management decision making (Payne & Bettman, 2004). The following quotes provide example of how students were considering their own personal values in ranking these criteria.

- "- I take a lot of pride in having a team that trusts me and have a mutual respect for each other, so reputation means a lot to me."
- "The order I selected above reflects my moral values and what I believe is just."
- "Productivity is the next important because it protects you and the rest of the employees' livelihoods."

Student Explain Rankings by Connecting to Other Criteria. Oftentimes, students would not describe the inherent value in a criteria, but instead, they would link one criteria's importance to that of another criteria as a form of justification. This approach includes the idea that prioritizing one criteria may "domino" into obtaining positive outcomes among other desirable criteria. Alternatively, linking criteria together may have been a method of explaining low ranking criteria or outcomes they described as "undesirable". Students who linked criteria may recognize the relationships among criteria, which may suggest a nuanced understanding of how competing criteria interact with their decision making. The following reflections provide example of how students would connect criteria to explain their importance.

- "Safety is always first because safety disasters can affect all of the other factors…"
- "Safety and productivity are the most important because that is what will keep the plant running and making money. If those are prioritized, the rest will follow."
- "I feel as though plant productivity and time management go hand in hand as finishing a project in a timely manner depends on how productive a plant is."

Student Explain Rankings with the Goal of Avoiding Potentially Negative Outcomes. Students also ranked criteria based on its contribution to avoiding a negative outcome. Frequently, this was associated with the safety criteria where students went into detail as to what they were avoiding by prioritizing safety or other criteria. The following quotes exemplify how students explained their rankings while considering a criteria's contribution to potential consequences.

- "Safety is important to prevent harm to your team."
- "Safety is the most important factor because an unsafe environment can cause unnecessary risk, injury, death, lawsuits, or other events that should be avoided at all costs."
- "Safety is the most important since the consequence may be severe."

- "Safety was my highest priority because I didn't want anyone getting hurt… Plant productivity, budget and time weren't my biggest concerns when the hurricane was coming through. I wanted to make sure large damages to the plant weren't going to affect production later after the storm had passed."

Student Explain Rankings with the Goal of Supporting Business Responsibilities. Students also described the importance of criteria based on their contribution to supporting business responsibilities. Business responsibilities include how students believed they should act as engineers in the workplace, what they believe their employer or supervisor would want from them, or their responsibility to company finances or productivity. The following reflection overwhelmingly utilizes business responsibilities to explain their criteria rankings.

- "Safety is next because if the staff is injured no one can work. Productivity is next because it is the same reason we have managers, operators, and inspectors: to get the job done. Budget is important next because you want to make sure you're making more than you are costing the company. Last is personal relationships because the main goal is to get the job done, not to make friends..."

Student Explain Rankings by Recalling Experience. Students drew from their personal experiences to explain their rankings primarily during their post-game reflection. The following two quotes show a student reflecting in the pre-game survey, where they reflected on personal experiences, and in the post-game survey, where they reflected on their Contents Under Pressure game play.

- Pre-game reflection: "Safety should always be the top priority when working in a chemical plant environment and I feel that is pretty self-explanatory. Budget comes in at number two because, as my parents always say, money doesn't grow on trees. Overshoot the budget too many times and people will stop giving you contracts..."
- Post-game reflection: "Safety was my top priority throughout the game as I think this is the top determining factor when making decisions. It was difficult to maintain balance between the three categories in the game, but I tried to focus on plant productivity after safety. Personal relationships and time usually came last."

Changes in Students' Criteria Ranking Explanation. The author tracked changes in themes between the pre- and post-game reflections to show potential development in how students explained their rankings. Overall, changes are shown using a bar chart in Figure 5.2. This figure shows that 18 fewer themes were found in the post-game reflections than in the pre-game reflections. While multiple themes could be coded into a single reflection, post-game reflections were shorter in length; pre-game reflections had an average of 105 words whereas post-game reflections had an average 68 words. Shorter reflections may contribute to fewer found themes in the post-game reflections.

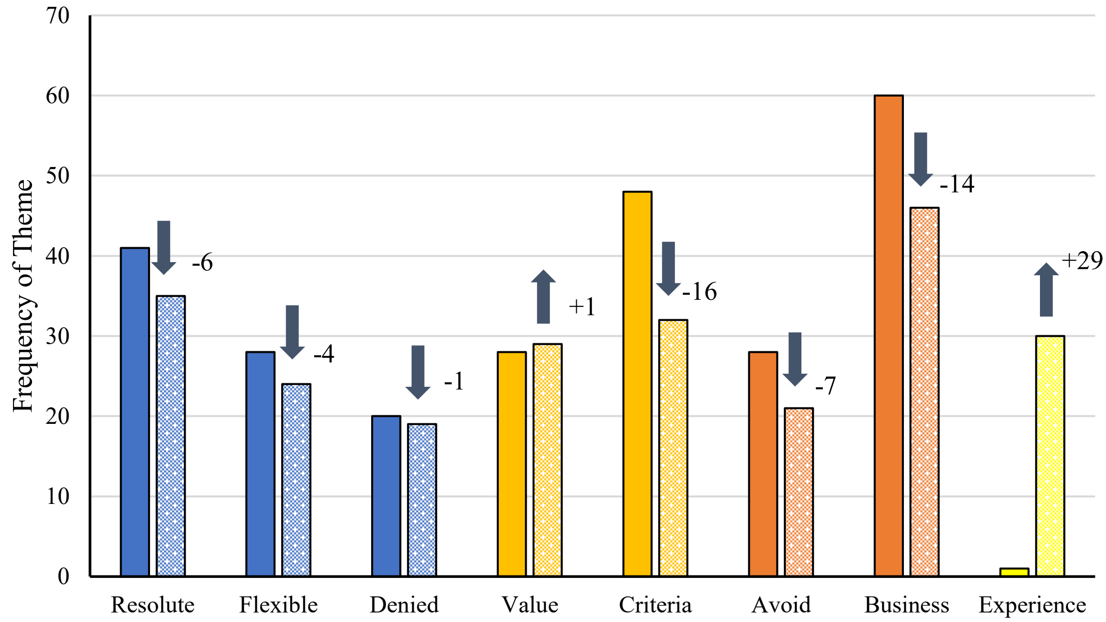


Figure 5.2. Frequency of themes in data from both pre- and post-game reflections (n=83).

The Connecting Criteria theme has noticeably fewer occurrences in these students' post-game reflections. Connections among concepts typically suggest complex understanding of a topic or of the relationships required to establish the connections among concepts (Ambrose et al., 2010; West et al., 2000). Yet, after engaging in a process safety game where they were given the opportunity to gain more complex understanding around process safety criteria, fewer students explained criteria rankings through their connections to other criteria. Alternatively, students may have gained a more complex understanding around criteria, forming connections among them, but potentially faced difficulty in transcribing that understanding into their reflection.

In addition, the Support Business Responsibilities theme had 14 fewer occurrences in the post-game reflection, yet this theme is still one of the most prevalent in this data. This result may be because of how deeply STEM education has been commoditized. For example, economic contribution and economic competitiveness has been embedded in historical decisions around how engineers are educated (Akera, 2017). Moreover, engineers and engineering education have devalued social engagement in favor of responsibilities to economic output (Carrigan & Bardini, 2021). This perspective is evident in students' ranking explanations as exemplified in the following quotes.

- "-because the only purpose of the plant is to produce company products for consumers."
- "-the sole purpose of your job is to make as much money for the plant as possible…"
- "-the main purpose of the chemical plant is to make some product and distribute it to their buyers, so the plant must be making product in an effective and efficient way."

While the Connecting Criteria and the Supporting Business Responsibility themes had *fewer* occurrences in the post-game reflection, the Experience theme had 29 *more* occurrences. It is likely during the pre-course reflection (at the beginning of their process safety course and prior to playing Contents Under Pressure) that students lacked relevant lived experiences to call on when explaining their rankings. For example, one student's set of reflections shows how they began to incorporate their lived experiences from Contents Under Pressure in their explanations.

- Pre-game reflection: “-And lastly, I selected personal relationships because although the decision to have a friendly work environment is important, it is also work.
- Post-game reflection: “-I ranked personal relationships third because as seen in the game if people were unhappy, it definitely affected productivity and also trust.”

Decision making literature specifies that experiences provide a pathway to making strategies for future decisions (Payne & Bettman, 2004). In addition, earlier arguments for Contents Under Pressure included providing students with a practice environment where they could build decision making experience (Barab & Duffy, 2012). Thus, it is possible students leveraged their recent experiences in a process safety course and in Contents Under Pressure to begin explaining their rankings during the post-game reflection.

The context of the survey likely influenced these responses. Students were told to *“Rank / order the items listed below based on how important they were to your decisions within Contents Under Pressure*,” and to “*Provide an explanation for the rank / order you have selected above and how this prioritization will impact your approach to decisions within a chemical plant environment.*” These prompts likely influenced students to reflect with their experiences from Contents Under Pressure. As such, while students may have gained lived experiences from which they can pull on to guide their future decisions, these results cannot distinctively claim that they will continue to leverage their lived experiences in future decisions that are farther removed from Contents Under Pressure. However, these reflections provide insight on how senior chemical engineering students are making decisions, which provides a basis for enhancements to process safety decision making curricula.

Research Question 5.3

This sub-research question asked, "*how do senior chemical engineering students' decisions within a digital process safety game indicate their prioritization of process safety criteria?"* This question is answered through a time series analysis on average metric scores in Contents Under Pressure. A time-series plot is shown in Figure 5.3, where the mean metric scores are plotted with standard deviation banding.

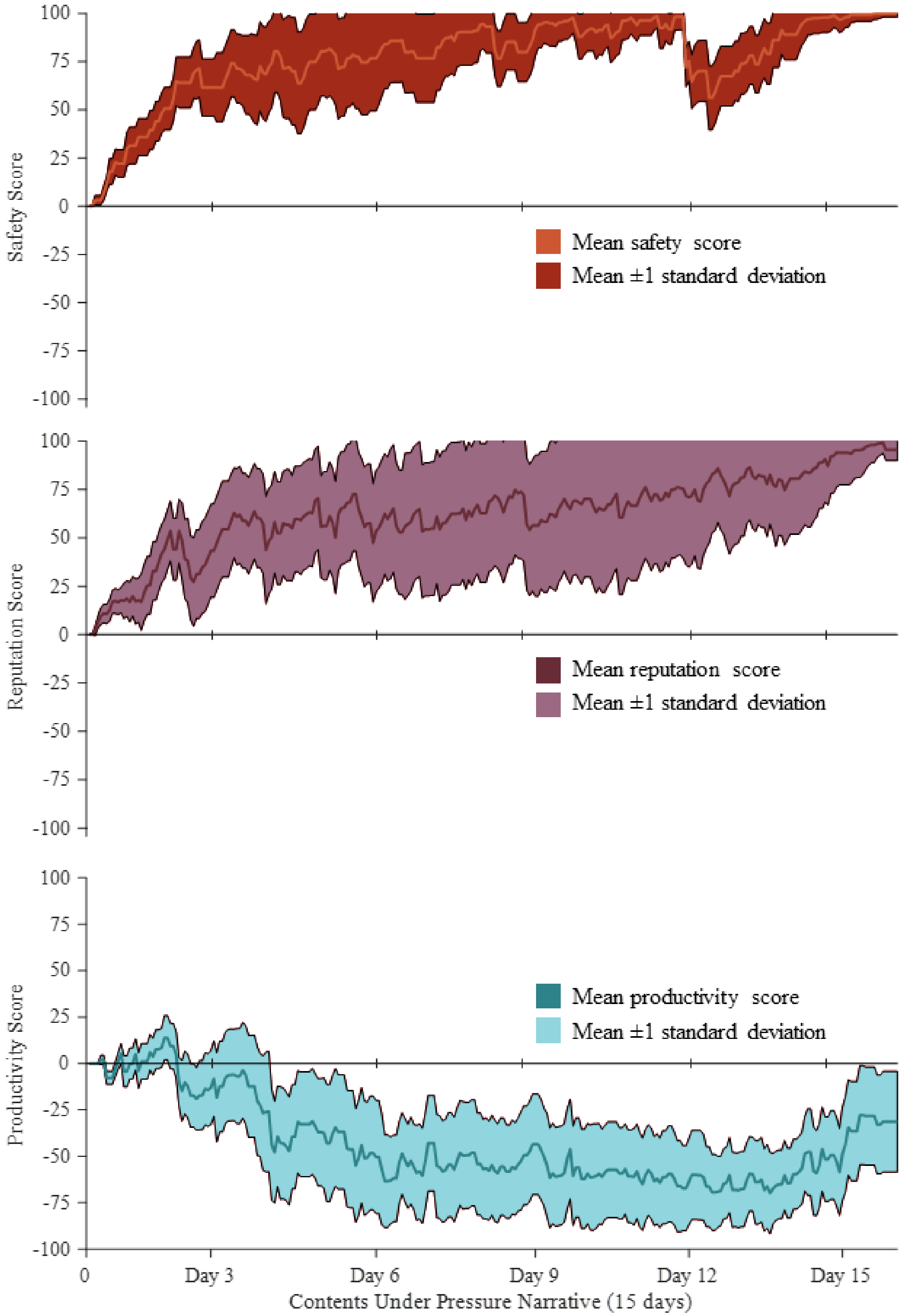


Figure 5.3. Time-series plot of game metrics in Contents Under Pressure (n=83).

Two points must be made ahead of discussing the results. The safety and reputation means and standard deviation banding flatline on the plot where the y-axis is equal to (100) exists because it is impossible to obtain a metric score greater than (100) in the game. Flatlining does not occur on the other extreme, where the y-axis is equal to (-100), because when a player obtains this score, their score is reset to (0). Secondly, the plot shows some extreme jumps in metric scores, such as the drop in reputation on Day 8, the drop in safety on Day 12, and the rise in plant output on Day 15. These trends do not reflect the decisions of players but changes in scores inherent to the narrative. On Day 8, players are confronted by their supervisor regarding a feud among the players' employees, "*Hey there. Got some time to chat? How are you feeling about your team?*" Regardless of decisions in the game, the feud will emerge, and the reputation score will drop after this interaction. Around Day 12 in the game, a large hurricane makes landfall over the chemical plant, and in order to instill a sense of urgency, the game drops players' safety scores. Similarly, the hurricane ends around Day 15, and the plant output score rises to signal the plant can return to normal operations.

Safety. Throughout the game, the safety metric is the most prioritized metric based on its highly scored mean and tight standard deviation banding. The banding suggests little variance in this metric among the players. The safety prioritization is frequently made by trading off with other criteria; for example, on Day 2, an employee (Emily) reports, "*Hey boss, I saw a minor spill. Cleaned it up but I noticed that it brought us close to our monthly limit,*" to which players can respond with, *"If we're that close, let's pass on reporting it,"* or "*We should file a report just to cover all our bases.*" While not claiming that players should takeaway that they should not report minor leaks, reporting the leak may take Emily from output objectives, which highlights the competing criteria players are forced to face.

The mean safety plot shows a drop during Day 8. An employee (Charles) requests, "*Heads up, Chief! I'd like to leave a couple hours early today for a doctor's appointment. Is that okay?*" Players may choose to let Charles leave and rely on others in the team or deny the leave request and claim the appointment must be made after work hours. Denying this request may take a toll on Charles' wellbeing. While the demand for medical leave is relevant in current events (BBC, 2022), players' decisions may have been influenced by piling deliverables delayed by recent storms. A decision choice in a previous prompt reminds players of their workload, "*There is still a lot of work to do after the past few days of rain. Maybe next week*." In addition, previous decisions that day to go to lunch with another employee, which would also slow down output, may have pressured players to reject further production slowdowns that day, "*Hey, do you want to grab lunch? I was wondering if I could ask some questions about how you got where you are.*"

Reputation. Players prioritized reputation as the second most important metric. While the mean score was high, this metric had much broader standard deviation banding. The broader standard deviation banding suggests there was greater variance in how players prioritized this metric. Example scenarios where the standard deviation banding decreased include feuds between game characters Victor Cob and Emily Nevil.

Victor, who is considered the *difficult employee* of the plant, says, "*Also, Wanda asked me how Emily was doing. I was honest - she seems to want to change things before she even knows how we do them. If she asks you about it, will you back me up?*" and "*I heard Emily already wants to take off for some conference. Did you tell her that she needs to earn her days off?*" The tightening standard deviation in this metric over these scenarios suggests that players may have a more consistent approach on how to respond to disagreements among employees when compared with other reputation based decisions

Plant Output. The TSA plot showed plant output was the least prioritized metric as the mean was a negative score for the majority of the narrative. While the mean score remained above (-100), this was the only metric that players "failed." Player failures most frequently occurred on Day 12 where 76 failures occurred among the 82 players (Figure 5.4). Day 12 is in the midst of a large hurricane making landfall over the plant where a worker is off site, players' family members call in with concern over the storm, and a safety supervisor is stranded in flood water. While the context of this day likely makes balancing the game metrics difficult, earlier explorative studies into the EPSRI reported that senior chemical engineering students had difficulty in making decisions around a similar hurricane dilemma (Stransky et al., 2020). This earlier study explained students' observed challenge with reference to unfamiliarity with certain ethical frameworks; however, this explanation may be extended through this present observation.

Students may find this game day to challenge their decision making because of their lack of experience. Students' prior practice in decision making often rely on case-studies and hypothetical situations where they are far removed from risks (Safety and Chemical Engineering Education, 2022); these experiences may provide students with time to rationalize their answer. However, Contents Under Pressure attempts to immerse students in the stress of these situations, where thoughtful deliberation may be difficult. This finding may identify a gap within process safety education: decision making in high stress situations. Hurricane and disaster type situations where engineers must make critical decisions are realistic (CSB, 2018). **As such, engineering students should become familiar with making decisions in high-stress, time-sensitive situations**.

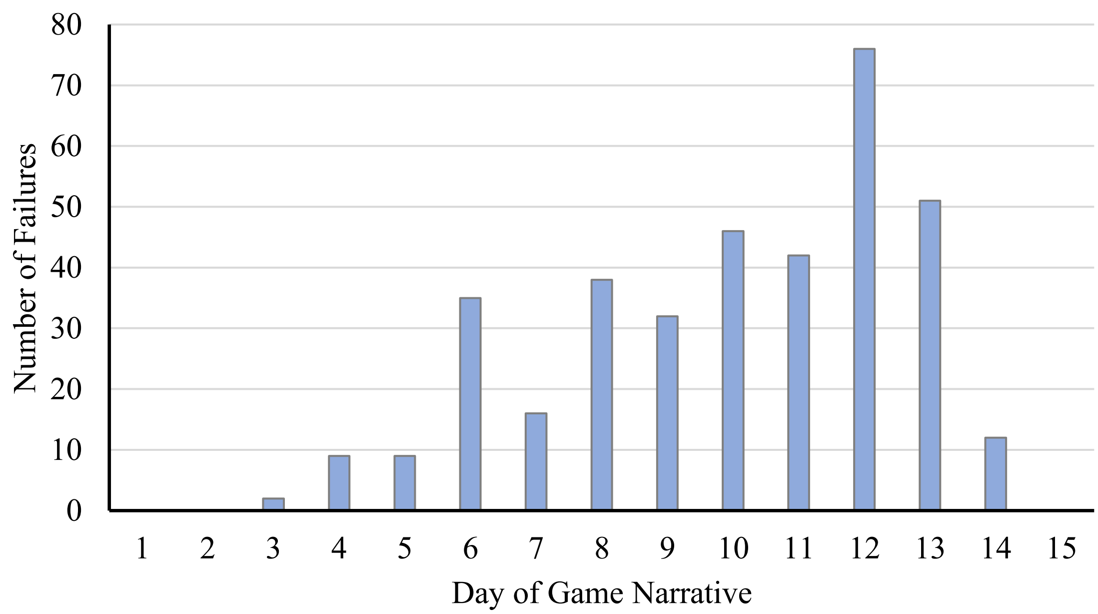


Figure 5.4. Bar chart showing number of failures per metric.

In answering the third research sub-question, senior chemical engineering students prioritized the safety metric in the Contents Under Pressure game. Safety was followed by reputation, yet this metric had broad variance in importance among players. Finally, plant output was the least prioritized metric and was the only metric that led to students' experiencing game failures. These findings may be limited by students' capacity to make decisions that align with their priorities. A majority of the decisions in Contents Under Pressure force players to make decisions with competing criteria; while students may desire to prioritize a game metric, inexperience with decision making or the game setting may hinder their ability to control desired outcomes. Moreover, these results are limited because these game metrics differ from process safety criteria. While these results speak to students' decision-making priorities, explicit claims regarding prioritizing criteria face transferability limitations. Instead, these results may provide evidence on students' priorities that may be contextualized with other forms of evidence to make robust claims.

Contextualizing Research Question 5

The overarching research question of this study asked, "*what role does competing criteria hold in the process safety decisions of engineering students?"* To answer this question, the author synthesizes the findings of the previous three research sub-questions by case-studying two students from the sample, Student A and Student B. The author randomly selected two students whose criteria rankings were within one standard deviation of the mean rankings reported in the first research sub-question. The similarities between these two students illustrate the consistency in how students ranked the process safety criteria. Student A's criteria rankings and explanations are shown in Figure 5.5. Student A recognizes that each of the five criteria are important to their decision making, yet initially, their focus is on “[making] money” over “emotions”. Student A’s reflection suggests they found personal relationships to be important to them, but they did not belong as a consideration in their decision making. After engaging with Contents Under Pressure, Student A changes the ranking of personal relationships from fifth to third and acknowledges that “amicable relationships” are required for a chemical plant to function.

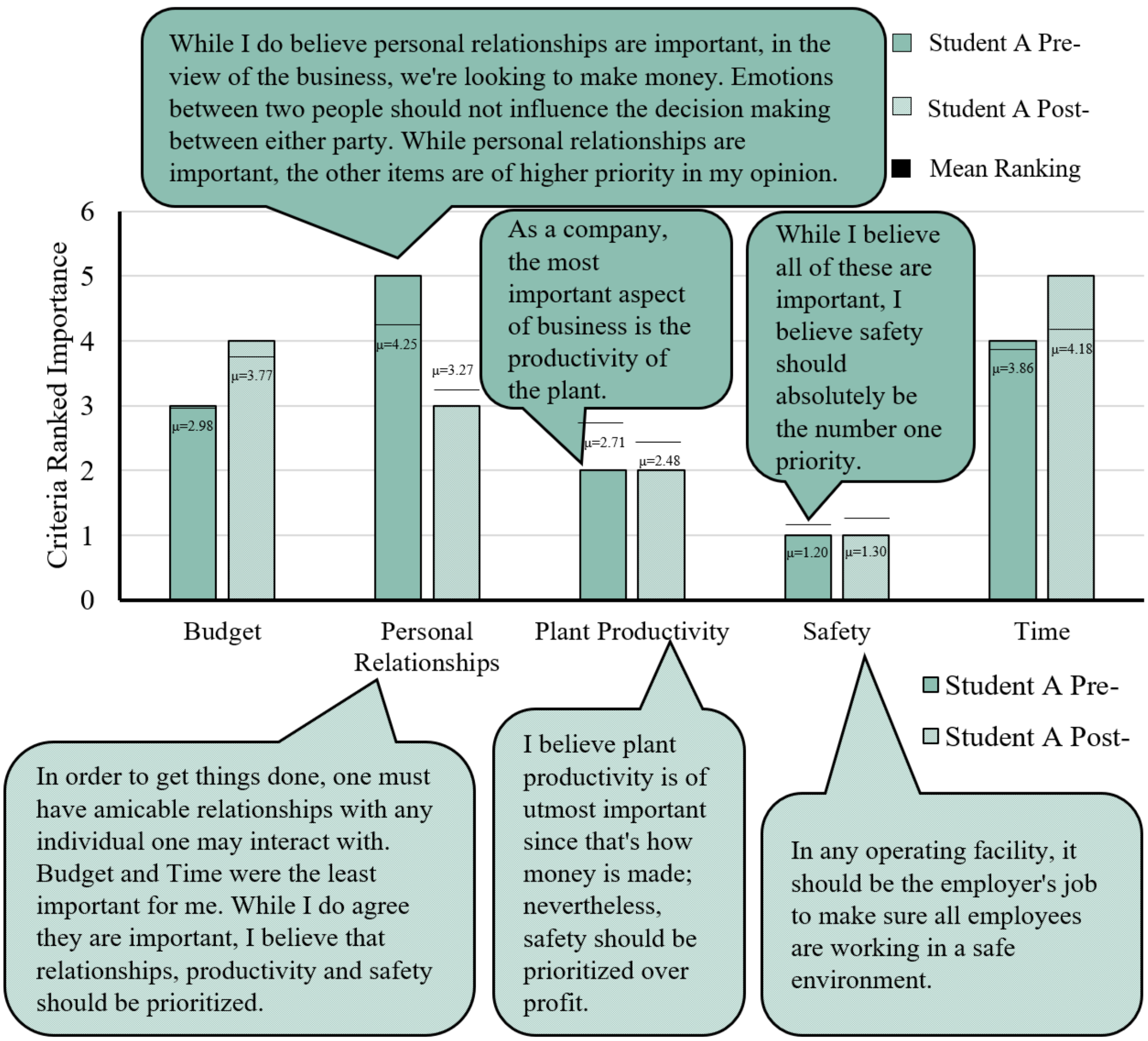


Figure 5.5. Student A’s rankings with reflection callouts from both pre- and post-game surveys.

Student B followed the trends of Student A as shown in Figure 5.6. Student B expresses that the safety criteria is of resolute importance, followed by other criteria. They rank personal relationships as fifth, yet they do not provide discernment why there is not "a big gap between these options." After engaging with Contents Under Pressure, Student B continues to rank safety of resolute importance, and they introduce a new theme of "running a successful plant" when explaining plant productivity. They also changed the ranking of personal relationships from fifth to third by expressing their personal value for "mental health", which may come from these relationships.

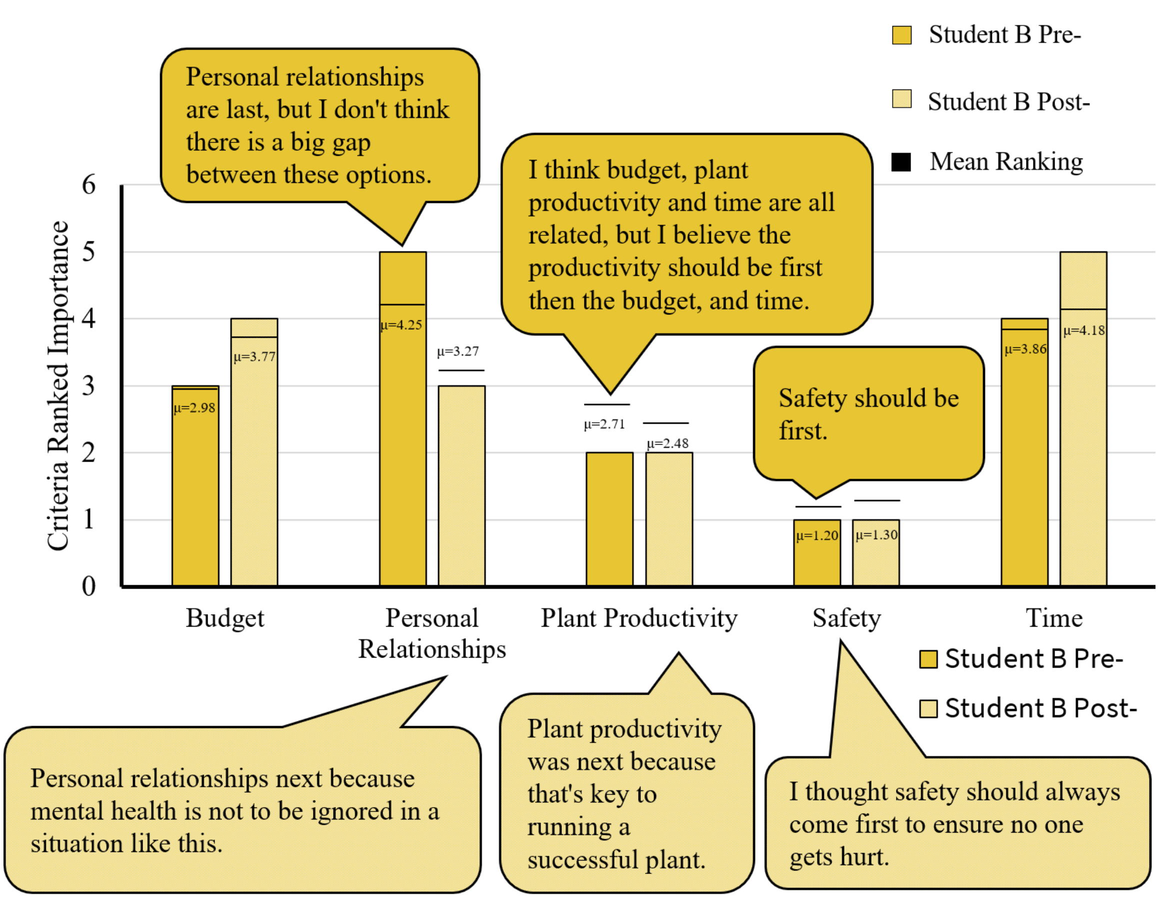


Figure 5.6. Student B's rankings with reflection callouts from both pre- and post-game surveys.

Both of these students claim to see value in all five of these criteria including personal relationships, yet they describe this criteria to just not measure up to the others. Both of these students' reflections were coded to be describing the importance of safety with explicit certainty. Despite both students making positive remarks about plant productivity in their pre- and post-game reflections, they experienced game failures in the plant output game metric Figure 5.7. This outcome reiterates how even though plant productivity was their second ranked criteria, students' may not have the capacity to control the state of the game metrics criteria. Capacity to manage over these criteria may be improved through additional practice (Payne & Bettman, 2004).

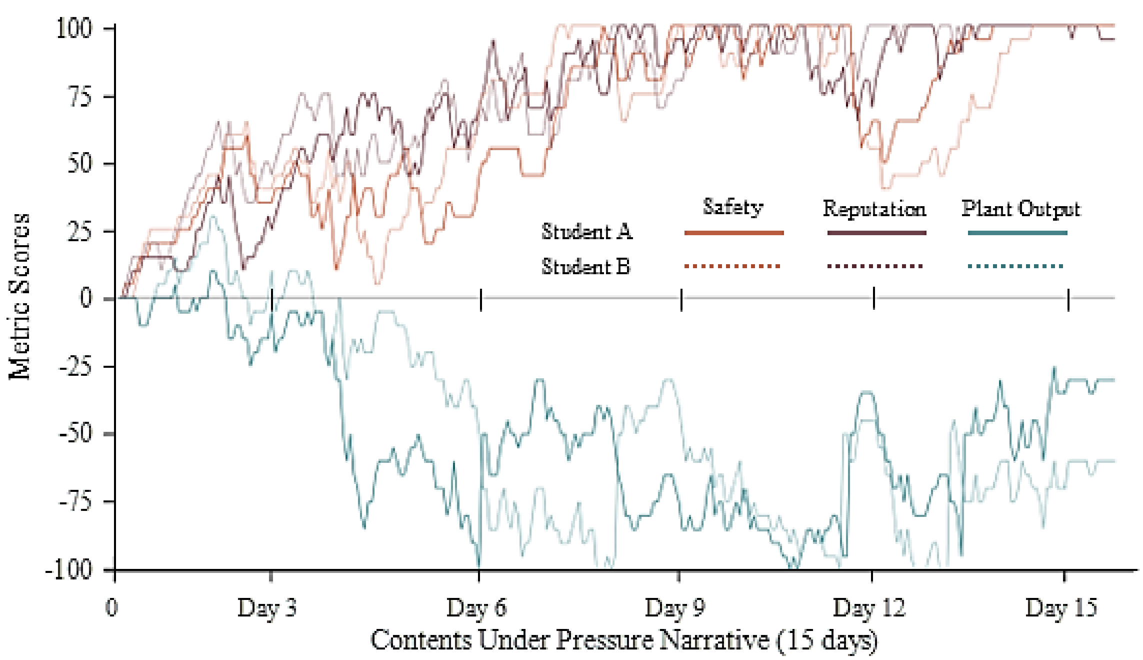


Figure 5.7. TSA plot of case study students' gameplay data.

Conclusions

The objective of this study is to gain insight on how chemical engineers consider process safety criteria in their decisions by studying senior chemical engineering students by answering the overarching research question, *what role does competing criteria hold in the process safety decisions of engineering students?* Throughout making decisions, engineers come across situations where criteria compete for priority, such as budget, productivity, personal relationships, safety, and time. Yet, issues with balancing these criteria by devaluing any one may lead to negative outcomes, despite intents for safety (Lehmann et al., 2022). For example, the competition among safety, plant productivity, and personal relationships have led to process incidents in the past (CSB, 2015, 2021). As such, this study targets understanding how engineering students consider these competing criteria and how their opinions about criteria evolve as engineering students build experience in making decisions with them. Empirical evidence of engineering students' criteria priorities is obtained through their decisions made in a digital process safety game. In addition, claimed criteria priorities were assessed through pre- and post-game surveys asking respondents to rank the five criteria in order of their importance and write an explanation of their ranking.

Student rankings found that safety was strongly prioritized, followed by plant productivity. When comparing students' pre-game rankings to their post-game rankings, personal relationships were increasingly prioritized from fifth to third in average importance. The results of an ANOVA showed a statistically significant change in ranking with a large effect. Whereas changes in plant productivity and safety were negligible to small.

Students initially explain their rankings by discussing the correlations among the criteria or by relating them to their responsibilities as an engineer working at a company. After playing the game, almost half of the students explained their rankings by recalling prior experiences, such as by prioritizing safety. In-game decisions show students' empirical priorities. Safety is prioritized above their reputation and plant output. While reputation was empirically prioritized over plant output, its exact importance varied greatly among students. Moreover, plant output was devalued by students until in-game operation "failures" occurred. To directly answer the overarching research question, engineering students consider competing criteria in their process safety decisions by prioritizing safety, often to the extent of discounting other criteria, and they justify these decisions by reflecting on responsibilities to businesses they work for or by reflecting on prior experiences.

This study found that students' perception of competing criteria can change after gaining experience in process safety decision making, such as by playing Contents Under Pressure as shown through the results of Research Sub-question 5.1. Engineers need to recognize the relevance of competing criteria on their decision making as issues with balancing these criteria can lead to process incidents. This study does not seek to instill an ordinally ranked set of criteria for engineers; instead, it draws attention to the relevance of competing criteria. Awareness and experience in making decisions with these competing criteria may be critical to avoiding further process incidents.

References

ABET. (2021a). *Accreditated Programs*. https://amspub.abet.org/aps/category-search?disciplines=31&disciplines=32&disciplines=33

ABET. (2021b). *Criteria for Accrediting Engineering Programs, 2021-2022*. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2021-2022/

Agency, E. C. (2023). *About ECHA*. https://echa.europa.eu/about-us

Akera, A. (2017). Setting the standards for engineering education: A history [scanning our past]. *Proceedings of the IEEE*, *105*(9), 1834–1843. https://doi.org/10.1109/JPROC.2017.2729406

Akinleye, D. D., McNutt, L. A., Lazariu, V., & McLaughlin, C. C. (2019). Correlation between hospital finances and quality and safety of patient care. *PLoS ONE*, *14*(8), 1–19. https://doi.org/10.1371/journal.pone.0219124

Bairaktarova, D., & Woodcock, A. (2017). Engineering Student's Ethical Awareness and Behavior: A New Motivational Model. *Science and Engineering Ethics*, *23*(4), 1129–1157. https://doi.org/10.1007/s11948-016-9814-x

Barab, S. A., & Duffy, T. (2012). From practice fields to communities of practice. In D. Jonassen & S. Land (Eds.), *Theoretical Foundations of Learning Environments* (2nd ed., pp. 25–55). Routledge.

Baybutt, P. (1996). *Human Factors in Process Safety and Risk Management: Needs for Models, Tools and Techniques*. https://www.primatech.com/images/docs/paper_human_factors_in_process_safety_and_risk_management_needs_for_models_tools_and_techniques.pdf

Baybutt, P. (2017a). Process safety incidents , cognitive biases and critical thinking. *Hydrocarbon Processing*, *2017*(April).

Baybutt, P. (2017b). The validity of engineering judgment and expert opinion in hazard and risk analysis: The influence of cognitive biases. *Process Safety Progress*, *37*(2), 205–210. https://doi.org/10.1002/prs.11906

Baybutt, P. (2015). The psychology of decision making in process hazard analysis. *Process Safety Spotlights 2015 - Topical Conference at the 2015 AIChE Spring Meeting and 11th Global Congress on Process Safety*, 448–465.

Bazerman, M. H., & Tenbrunsel, A. (2011). *Blind Spots: Why We Fail to Do What's Right and What to Do about It*. Princeton University Press.

British Broadcasting Corporation (BBC). (2022, December 7). When are the train strikes and what is the dispute about ? *BBC*. https://www.bbc.com/news/business-61634959

Biddle, E., & Afanuh, S. (2015). *Supporting Prevention through Design (PtD) Using Business Value Concepts*. https://www.cdc.gov/niosh/docs/wp-solutions/2015-198/pdfs/2015-198.pdf?id=10.26616/NIOSHPUB2015198

Bland, J. M., & Altman, D. G. (1995). Multiple Signifivance Tests: the Bonferroni method. *The British Medical Journal*, *310*(170). https://www.bmj.com/content/310/6973/170

Bodnar, C., Anastasio, D., Enszer, J. A., & Burkey, D. D. (2016). Engineers at play: Games as teaching tools for undergraduate engineering students. *Journal of Engineering Education*, *105*(1), 147–200.

Bodnar, C., Dringenberg, E., Butler, B., Burkey, D. D., Anastasio, D. D., & Cooper, M. (2020). Revealing the decision-making processes of chemical engineering students in process safety contexts. *Chemical Engineering Education*, *54*(1), 22–30.

Borenstein, J., Drake, M. J., Kirkman, R., & Swann, J. L. (2010). The Engineering and Science Issues Test (ESIT): A Discipline-Specific Approach to Assessing Moral Judgment. *Science and Engineering Ethics*, *16*(2), 387–407. https://doi.org/10.1007/s11948-009-9148-z

Bormann, D., & Greitemeyer, T. (2015). Immersed in Virtual Worlds and Minds: Effects of In-Game Storytelling on Immersion, Need Satisfaction, and Affective Theory of Mind. *Social Psychological and Personality Science*, *6*(6), 646–652. https://doi.org/10.1177/1948550615578177

Burkey, D. D., Anastasio, D. D., Bodnar, C., & Cooper, M. (2020). *Collaborative Research: Experiential Process Safety Training for Chemical Engineers*. STEM for All Video Showcase. https://stemforall2020.videohall.com/presentations/1691

Butler, B. (2018). *The Creation, Validation, and Implementation of the Engineering Process Safety Research Instrument* [Rowan University]. rdw.rowan.edu/etd/2627

Butler, B., Anastasio, D. D., Burkey, D. D., Cooper, M., & Bodnar, C. (2018). Work in Progress: Content Validation of an Engineering Process Safety Decision- making Instrument (EPSRI). *2018 ASEE Annual Conference & Exposition*. https://peer.asee.org/31279

Butler, B., Bodnar, C., Cooper, M., Burkey, D. D., & Anastasio, D. D. (2019). Towards understanding the moral reasoning process of senior chemical engineering students in process safety contexts. *Education for Chemical Engineers*, *28*, 1–12. https://doi.org/10.1016/j.ece.2019.03.004

Carrigan, C., & Bardini, M. (2021). Majorism: Neoliberalism in Student Culture. *Anthropology & Education Quarterly*, *52*(1), 42–62. https://doi.org/101111/aeq.12361

Carter, M., Gibbs, M., & Harrop, M. (2012). Metagames, paragames and orthogames: A new vocabulary. *Foundations of Digital Games 2012, FDG 2012 - Conference Program*, 11–17. https://doi.org/10.1145/2282338.2282346

CCPS, C. for C. P. S. (2022). *Safety and Chemical Engineering Education (SAChE) Certificate Program*. https://www.aiche.org/ccps/education/safety-and-chemical-engineering-education-sache-certificate-program

Charmaz, K. (2006). *Constructing Grounded Theory: A practical Guide Through Qualitative Analysis*. Sage Publications. https://doi.org/10.1186/s12868-016-0320-5

Clancy, R. F. (2020a). Ethical reasoning and moral foundations among engineering students in China. *ASEE's Virtual Conference: At Home with Engineering Education*.

Clancy, R. F. (2020b). The Ethical Education and Perspectives of Chinese Engineering Students: A Preliminary Investigation and Recommendations. In *Science and Engineering Ethics* (Vol. 26, Issue 4). Springer Netherlands. https://doi.org/10.1007/s11948-019-00108-0

Clancy, R. F. (2021). The Relations between Ethical Reasoning and Moral Intuitions among Engineering Students in China. *ASEE Annual Conference and Exposition, Conference Proceedings*.

Clancy, R. F., Sessford, J. R., An, L., & Ge, Y. (2017). Which factors are correlated with engineering students' expectations of ethical issues? *ASEE Annual Conference and Exposition, Conference Proceedings*, *2017-June*. https://doi.org/10.18260/1-2--29124

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates. http://utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf

Cong, G. P., Shi, X., & Meng, T. Y. (2016). HAZOP-LOPA-Based Corrosion Risk Identification and Control. *Applied Mechanics and Materials*, *853*, 449–452. https://doi.org/10.4028/www.scientific.net/amm.853.449

Cortina, J. M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Crawford, C. (1997). *The Art of Computer Game Design*. Washington State University. https://www.digitpress.com/library/books/book_art_of_computer_game_design.pdf

Crawford, C. (2003). *Chris Crawford on game design*. New Riders.

Creswell, J. W. (2009). The Use of Theory. In *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed., pp. 49–72). Sage Publications.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*(3), 297–334.

Crowl, D. A., & Louvar, J. F. (2019). *Chemical Process Safety: Fundamentals with Applications* (Fourth). Pearson Education, Inc.

CSB (United States Chemical Saefty and Hazard Investigation Board). (2009). Investigation Report: T2 Laboratories, INC. In *U. S. Chemical Safety and Hazard Investigation Board* (Issue 2008).

CSB (United States Chemical Saefty and Hazard Investigation Board). (2011). *Investigation Report: E.I. DuPont de Nemours & Co., Inc*.

CSB (United States Chemical Saefty and Hazard Investigation Board). (2014). *Regulatory Report: Chevron Richmond Refinery Pipe Ruture and Fire* (Issue January). https://doi.org/10.32388/u2vkmu

CSB (United States Chemical Saefty and Hazard Investigation Board). (2015). *Final Investigation Report: Chevron Richmond Refinery #4 Crude Unit*. https://www.csb.gov/chevron-refinery-fire/

CSB (United States Chemical Saefty and Hazard Investigation Board). (2017). *Key Lessons from the ExxonMobil Baton Rouge Refinery Isobutane Release and Fire*.

CSB (United States Chemical Saefty and Hazard Investigation Board). (2018). *Investigation Report: Non-Condensable Gas System Explosion at PCA DeRidder Paper Mill*. *April 2018*, 102.

CSB (United States Chemical Saefty and Hazard Investigation Board). (2019a). *Factual Update: Fire and Explosions at Philadelphia Energy Solutions Refinery Hydrofluoric Acid Alkylation Unit*.

CSB (United States Chemical Saefty and Hazard Investigation Board). (2019b). *Investigation Report: Gas Well Blowout and Fire at Pryor Trust Well 1H-9*.

CSB (United States Chemical Saefty and Hazard Investigation Board). (2019c). *Investigation Report: Toxic Chemical Release at the DuPont La Porte Chemical Facility*.

CSB (United States Chemical Saefty and Hazard Investigation Board). (2021). *Investigation Report: Hydrogen Sulfide Release at Aghorn Operating Waterflood Station* (Issue May).

Derryberry, W. P., & Thoma, S. J. (2005). Functional differences: Comparing moral judgement developmental phases of consolidation and transition. *Journal of Moral Education*, *34*(1), 89–106. https://doi.org/10.1080/03057240500049372

DeVellis, R. F. (2016). *Scale Development Theory and Applications* (4th ed., Vol. 4). Sage Publications.

Dönmez, K., & Uslu, S. (2020). The effect of management practices on aircraft incidents. *Journal of Air Transport Management*, *84*. https://doi.org/10.1016/j.jairtraman.2020.101784

Dörner, R., Göbel, S., Effelsberg, W., & Wiemeyer, J. (2001). Introduction. In R. Dörner, S. Göbel, W. Effelsberg, & J. Wiemeyer (Eds.), *Serious Games: Foundations, Concepts, and Practices*. Springer International Publishing.

Dörner, R., Göbel, S., Effelsberg, W., & Wiemeyer, J. (2016). *Serious games: Foundations, Concepts and Practices* (1st ed.). Springer International Publishing. https://doi.org/10.7146/pas.v33i80.111719

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, *37*(2), 90–93. https://doi.org/10.1037/h0058073

Encinosa, W. E., & Bernard, D. M. (2005). Hospital finances and patient safety outcomes. *Inquiry*, *42*(1), 60–72. https://doi.org/10.5034/inquiryjrnl_42.1.60

Floyd, F. J., & Widaman, K. F. (1995). Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment*, *7*(3), 286–299.

Friedman, S., & Greenhaus, J. (2011). Choosing Work or Family ... or Both? In *Work and Family-Allies or Enemies?: What Happens When Business Professionals Confront Life Choices* (Issue 1, pp. 19–40). Oxford Scholarship Online. https://doi.org/10.1093/acprof:oso/9780195112757.001.0001

Geddes, K. (2019, October 28). Texas couple dies after exposure to H_2S gas. *KENS5*. https://www.kens5.com/article/news/local/husband-wife-die-h2s-gas-exposure/513-ce8f684b-6c99-46d0-880f-5a7e0a131402

Ghorbani, M., Maciejewski, A. A., Siller, T. J., Chong, E. K. P., Omur-Ozbek, P., & Atadero, R. A. (2018). Incorporating ethics education into an electrical and computer engineering undergraduate program. *ASEE Annual Conference and Exposition, Conference Proceedings*, *2018-June*. https://doi.org/10.18260/1-2--30645

Gilligan, C., & Attanucci, J. (1988). Two Moral Orientations: Gender Differences and Similarities. *Merrill-Palmer Quarterly*, *34*(3), 223–237. https://www.jstor.org/stable/23086381

Griethuijsen, R. van, Eijck, M. W. van, Haste, H., den Brok, P. J., Skinner, N. C., Mansour, N., Gencer, A. S., & BouJaoude, S. (2015). Global patterns in students' views of science and interest in science. *Research in Science Education*, *45*(4), 581–603. https://doi.org/10.1007/s11165-014-9438-6

Hair, Anderson, Tatham, & Black. (1998). *Multivariate Data Analysis* (5th ed.).

Harris, M. A., Peck, R. F., Colton, S., Morris, J., Neto, E. C., & Kallio, J. (2009). A Combination of hand-held models and computer imaging programs helps students answer oral questions about molecular structure and function: A controlled investigation of student learning. *CBE Life Sciences Education*, *8*(1), 29–43. https://doi.org/10.1187/cbe.08-07-0039

Hejdenberg, A. (2005). *The Psychology Behind Games*. Game Developer. https://www.gamedeveloper.com/design/the-psychology-behind-games

Hendrickson, C., & Au, T. (2008). Quality Control and Safety During Construction. In C. Hendrickson (Ed.), *Project Management for Construction: Fundamental Concepts for Owners, Engineers, Architects, and Builders*. Prentice Hall. https://www.cmu.edu/cee/projects/PMbook/13_Quality_Control_and_Safety_During_Construction.html

Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for use in Survey Questionnaires. *Organizational Research Methods*, *1*(1), 104–121.

Howard, R. A., & Abbas, A. E. (2016). Foundations of Decision Analysis. In *Handbook of Decision Analysis*. Pearson.

Johnsen, H. M., Fossum, M., Vivekananda-Schmidt, P., Fruhling, A., & Slettebø, Å. (2018). Nursing students' perceptions of a video-based serious game's educational value: A pilot study. *Nurse Education Today*, *62*(May 2017), 62–68. https://doi.org/10.1016/j.nedt.2017.12.022

Joiner, R., Iacovides, J., Owen, M., Gavin, C., Clibbery, S., Darling, J., & Drew, B. (2011). Digital Games, Gender and Learning in Engineering: Do Females Benefit as Much as Males? *Journal of Science Education and Technology*, *20*(2), 178–185. https://doi.org/10.1007/s10956-010-9244-5

Juego-Studio. (2023). *SIMULATION INDUSTRIAL TRAINING*. https://www.juegostudio.com/simulation-industrial-training-case-study

Kapp, K. M. (2012). *The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education*. Pfeiffer.

Khammar, A., Yarahmadi, M., & Madadizadeh, F. (2020). What is analysis of covariance (ANCOVA) and how to correctly report its results in medical research? *Iranian Journal of Public Health*, *49*(5), 1016–1017. https://doi.org/10.18502/ijph.v49i5.3227

Kline, R. B. (2015). Principles and practices of structural equation modelling. In *Methodology in the social sciences* (4th ed.). The Guilford Press.

Kohlberg, L. (1981). *Essays on Moral Development.* Jossey-Bass.

Kohlberg, L. (1985). Resolving moral conflicts within the just community. In C. B. Harding (Ed.), *Moral dilemmas: philosophical and psychological issues in the development of moral reasoning* (pp. 71–97). Precedent Publishing.

Kohlberg, L., & Hersh, R. H. (1977). Moral Development: A Review of the Theory. *Theory Into Practice*, *16*(2), 53–59. https://doi.org/10.1080/00405847709542675

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Lehmann, S., Bartlett, A., Menon, A., Crosman, J., Studd, A., & Dubey, K. (2022). *SPHERA INSIGHTS : Safety Report 2022*. https://sphera.com/safety-report-2022-report-thank-you/?utm_source=aiche&utm_medium=syndication

Lichtenstein, S., Lichtenstein, G., & Higgs, M. (2017). Personal values at work: A mixed-methods study of executives' strategic decision-making. *Journal of General Management*, *43*(1), 15–23. https://doi.org/10.1177/0306307017719702

Lietz, C. A., Langer, C. L., & Furman, R. (2006). Establishing Trustworthiness in Qualitative Research in Social Work: Implications from a Study Regarding Spirituality. *Qualitative Social Work*, *5*(4), 441–458. https://doi.org/10.1177/1473325006070288

Ma, M., & Oikonomou, A. (2017). Serious games and edutainment applications: Volume II. In *Serious Games and Edutainment Applications: Volume II: Vol. II*. https://doi.org/10.1007/978-3-319-51645-5

Ma, M., Oikonomou, A., & Jain, L. C. (Eds.). (2011). *Serious Games and Edutainment Applications*. Springer.

Mandal, J., Ponnambath, D. K., & Parija, S. C. (2016). Utilitarian and deontological ethics in medicine. *Trop Parasitol*, *6*(1), 5–7. https://doi.org/10.4103%2F2229-5070.175024

Martínez-durá, R. J., Arevalillo-herráez, M., García-fernández, I., Gamón-giménez, M. A., & Rodríguez-cerro, A. (2011). Serious Games and Edutainment Applications. In M. Ma, An. Oikonomou, & L. C. Jain (Eds.), *Serious Games and Edutainment Applications* (pp. 107–124). Springer. https://doi.org/10.1007/978-1-4471-2161-9

Mayhew, M. J., Pascarella, E. T., Trolian, T., & Selznick, B. (2015). Measurements Matter: Taking the DIT-2 Multiple Times and College Students' Moral Reasoning Development. *Research in Higher Education*, *56*(4), 378–396. https://doi.org/10.1007/s11162-014-9348-5

McDaniel, C. N., Lister, B. C., Hanna, M. H., & Roy, H. (2007). Increased Learning Observed in Redesigned Introductory Biology Course that Employed Web-enhanced, Interactive Pedagogy. *CBE-Life Sciences Education*, *6*, 243–249. https://doi.org/10.1187/cbe.07

McDonald, A. J., & Hansen, J. R. (2009). *Truth, Lies, and O-Rings: Inside the Space Shuttle Challenger Disaster Paperback*. University Press of Florida.

Meyers, L. S., Gamst, G., & Guarino, A. (2017a). Principal Component Analysis and Exploratory Factor Analysis. In *Applied Multivariate Research: Design and Interpretation2* (3rd ed., pp. 404–444). Sage Publications.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2017b). *Applied Multivariate Research: Design and Interpretation* (3rd ed.). SAGE Publications.

Microsoft. (2013). *Language Quality Game - Player Instructions*. https://social.technet.microsoft.com/wiki/contents/articles/9301.language-quality-game-player-instructions.aspx

Mildner, P., & Mueller, F. (2016). Design of Serious Games. In R. Dörner, S. Göbel, W. Effelsberg, & J. Wiemeyer (Eds.), *Serious Games: Foundations, Concepts, and Practices* (pp. 57–82). Springer International Publishing.

Miles, M., Huberman, A. M., & Saldana, J. (2014). Fundamentals of Qualitative Data Analysis. In *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). Sage Publications.

Mkpat, E., Reniers, G., & Cozzani, V. (2018). Process safety education: A literature review. *Journal of Loss Prevention in the Process Industries*, *54*(February), 18–27. https://doi.org/10.1016/j.jlp.2018.02.003

Money, J. (2020, January 29). Pittsburg County jury holds mud company partly responsible for fatal blowout two years ago. *The Oklahoman*. https://www.oklahoman.com/story/business/columns/2020/01/29/pittsburg-county-jury-holds-mud-company-partly-responsible-for-fatal-blowout-two-years-ago/60370872007/

Murzi, H., Mazzurco, A., Pikaar, I., & Gibbes, B. (2019). Measuring development of environmental awareness and moral reasoning: A case-study of a civil engineering course. *European Journal of Engineering Education*, *44*(6), 954–968.

Nakamura, J., & Csikszentmihalyi, M. (2014). The Concept of Flow. In *The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi* (pp. 239–263). https://doi.org/10.1007/978-94-017-9088-8

Ness, A. (2015). Lessons learned from recent process safety incidents. *Chemical Engineering Progress*, 23–29. https://www.aiche.org/sites/default/files/cep/20150323.pdf

Newstetter, W. C., & Svinicki, M. D. (2014). Learning Theories for Engineering Education Practice. In A. Johri & B. M. Olds (Eds.), *Cambridge Handbook of Engineering Education Research* (pp. 29–46). Cambridge University Press.

Nicaise, M., Gibney, T., & Crane, M. (2000). Toward an Understanding of Authentic Learning: Student Perceptions of an Authentic Classroom. *Journal of Science Education and Technology*, *9*, 79–94. https://doi.org/10.1023/A:1009477008671

Orvis, K. A., Moore, J. C., Belanich, J., Murphy, J. S., & Horn, D. B. (2010). Are soldiers gamers? videogame usage among soldiers and implications for the effective use of serious videogames for military training. *Military Psychology*, *22*(2), 143–157. https://doi.org/10.1080/08995600903417225

Osberg, T. M., & Shrauger, J. S. (1986). Self-Prediction: Exploring the Parameters of Accuracy. *Journal of Personality and Social Psychology*, *51*(5), 1044–1057.

Osborne, J. (2012a). Being True to the Target Population : Debunking the Myth of Representativeness. In *Best Practices in Data Cleaning : A Complete Guide to Everything You Need to Do Before and After Collecting Your Data* (pp. 43–70). Sage Publications. https://doi.org/https://dx.doi.org/10.4135/9781452269948

Osborne, J. (2012b). Dealing with Missing or Incomplete Data : Debunking the Myth of Emptiness. In *Best Practices in Data Cleaning : A Complete Guide to Everything You Need to Do Before and After Collecting Your Data* (pp. 105–138). Sage Publications. https://doi.org/https://dx.doi.org/10.4135/9781452269948

Osborne, J. (2012c). Power and Planning for Data Collection : Debunking the Myth of Adequate Power. In *Best Practices in Data Cleaning : A Complete Guide to Everything You Need to Do Before and After Collecting Your Data* (pp. 19–42). Sage Publications. https://doi.org/https://dx.doi.org/10.4135/9781452269948

Pasin, F., & Giroux, H. (2011). The impact of a simulation game on operations management education. *Computers and Education*, *57*(1), 1240–1254. https://doi.org/10.1016/j.compedu.2010.12.006

Payne, J., & Bettman, J. (2004). Walking with the Scarecrow: The Information Processing Approach to Decision Research. In D. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making*. Blackwell Publishing.

Pearson, R. H., & Mundfrom, D. J. (2010). Recommended sample size for conducting exploratory factor analysis on dichotomous data. *Journal of Modern Applied Statistical Methods*, *9*(2), 359–368. https://doi.org/10.22237/jmasm/1288584240

Pett, M., Lackey, N., & Sullivan, J. (2011a). Assessing the Characteristics of Matrices. In *Making Sense of Factor Analysis* (pp. 50–84). SAGE Publications. https://doi.org/10.4135/9781412984898

Pett, M., Lackey, N., & Sullivan, J. (2011b). Interpreting Factors and Generating Factor Scores. In *Making Sense of Factor Analysis* (pp. 17–225). SAGE Publications. https://doi.org/10.4135/9781412984898

Pett, M., Lackey, N., & Sullivan, J. (2011c). Reporting and Replicating the Results. In *Making Sense of Factor Analysis* (pp. 226–240). SAGE Publications. https://doi.org/10.4135/9781412984898

Pett, M., Lackey, N., & Sullivan, J. (2011d). Rotating the Factors. In *Making Sense of Factor Analysis* (pp. 131–166). SAGE Publications. https://doi.org/10.4135/9781412984898

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A Motivational Model of Video Game Engagement. *Review of General Psychology*, *14*(2), 154–166. https://doi.org/10.1037/a0019440

Radinsky, J., Bouillion, L., Lento, E. M., & Gomez, L. M. (2001). Mutual benefit partnership: A curricular design for authenticity. *Journal of Curriculum Studies*, *33*(4), 405–430. https://doi.org/10.1080/00220270118862

Reed, J. (2022). *Investigating Game-Based Instruction as a Tool for Engineering Ethics Education in a First-Year Engineering Program* [Rowan University]. https://rdw.rowan.edu/etd/2969

Reed, J., Streiner, S., Burkey, D. D., Cimino, R., Pascal, J., & Young, M. (2021). "Mapping" the Landscape of First-Year Engineering Students' Conceptualizations of Ethical Decision Making. *2021 ASEE Annual Conference*.

Reilly, J. M., & Dede, C. (2019). Differences in student trajectories via filtered time series analysis in an immersive virtual world. *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, 130–134. https://doi.org/10.1145/3303772.3303832

Rest, J. (1974). Judging the important issues in moral dilemmas: An objective measure of development. *Developmental Psychology*, *10*(4), 491–501. https://doi.org/10.1037/h0036598

Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (1999). A Neo-Kohlbergian Approach: The DIT and Schema Theory. *Educational Psychology Review*, *11*(4), 291–324.

Rest, J., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). DIT2: Devising and Testing a Revised Instrument of Moral Judgement. *Journal of Educational Psychology*, *91*(4), 664–659.

Rest, J., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (2000). A Neo-Kohlbergian approach to morality research. *Journal of Moral Education*, *29*(4). https://doi.org/10.1080/713679390

Rest, J., Thoma, S., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology*, *89*(1), 5–28. https://doi.org/10.1037/0022-0663.89.1.5

Rest, J., Thoma, S. J., Narvaez, D., & Bebeau, M. J. (1997). Alchemy and Beyond: Indexing the Defining Issues Test. *Journal of Educational Psychology*, *89*(3), 498–507. https://doi.org/10.1037/0022-0663.89.3.498

Rockstar Games, Inc. (2013). *Grand Theft Auto V* (Steam Version). https://www.rockstargames.com/gta-v

Rogers, Wi. P., Armstrong, N., Acheson, D. C., Covert, E., Feynman, R., Holtz, R. B., Kutyna, D. J., Ride, S. K., Rummel, R. W., Sutter, J. F., Walker, A. B. C., Wheelon, A. D., Yeager, C. E., & Keel, A. G. (1986). *Space Shuttle Accident and the Rogers Commission Report*.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 347–363. https://doi.org/10.1007/s11031-006-9051-8

Safety and Chemical Engineering Education. (2022). *About*. http://www.sache.org/

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers* (3rd ed.). SAGE.

Saucier, Gerald, & Goldberg, L. R. (1996). Evidence for the Big Five in analyses of familiar English personality adjectives. *European Journal of Personality*, *10*(1), 61–77.

Saucier, Gerard, & Goldberg, L. R. (1996). The Language of Personality: Lexical Perspectives on the Five-Factor Model. In J. S. Wiggins (Ed.), *The Five-Factor Model of Personality: Theoretical Perspectives* (pp. 21–50). The Guilford Press.

Sawyer, R., Rowe, J., Azevedo, R., & Lester, J. (2018). Filtered time series analyses of student problem-solving behaviors in game-based learning. *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018*, 229–238.

Schrier, K. (2021). *We the Gamers: How Games Teach Ethics and Civics*. Oxford University Press.

Shafer-Landau, R. (2018). *The Fundamentals of Ethics* (4th ed.). Oxford University Press.

Siegler, R. S. (1997). Concepts and Methods for Studying Cognitive Change. In E. Amsel & K. A. Renninger (Eds.), *Change and development: Issues of theory, method, and application* (pp. 77–98). Lawrence Erlbaum Associates Publishers.

Stransky, J., Bassett, L., Bodnar, C., Anastasio, D. D., Burkey, D. D., & Cooper, M. (2021). Understanding Student Motivation to Engage in the Contents Under Pressure Digital Game. *ICL2021 – 24th International Conference on Interactive Collaborative Learning*, *September*, 1716–1727.

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—or Why the P Value Is Not Enough . *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/jgme-d-12-00156.1

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tawa, J. (2017). "Walk a Mile in My Shoes": A Virtual World Exercise for Fosterug Students' Subjective Understanding of the Experiences of People of Color. In M. Ma & A. Oikonomou (Eds.), *Serious Games and Edutainment Applications: Volume II*. Springer.

Tenbrunsel, A., Diekmann, K. A., Wade-Benzoni, K. A., & Bazerman, M. H. (2010). The ethical mirage: A temporal explanation as to why we are not as ethical as we think we are. *Research in Organizational Behavior*, *30*(2010), 153–173.

Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, *17*(2), 223–236. https://doi.org/10.1023/B:SORE.0000027411.35832.53

Testino, R. P. (2007). *Ethical Fading and Biased Assessments of Fairness* (Issue June). Linköpings Universitet.

Thoma, S. J., & Rest, J. R. (1999). The relationship between moral decision making and patterns of consolidation and transition in moral judgment development. *Developmental Psychology*, *35*(2), 323–334. https://doi.org/10.1037/0012-1649.35.2.323

Vaughen, B. (2012). An approach to help departments meet new abet process safety requirements. *Chemical Engineering Education*, *46*(2), 129–134.

Velicer, B. W. F., Hoeppner, B. B., & Goodwin, M. S. (2012). Time-Series Study. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 1520–1528). SAGE Publications, Inc. https://dx.doi.org/10.4135/9781412961288

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2016). Computer Gaming and Interactive Simulations for Learning: A Meta-Analysis. *Journal of Educational Computing Research*, *34*(3). https://doi.org/10.2190/FLHV-K4WA-WPVQ-H0YM

Walther, J., Sochacka, N. W., Benson, L. C., Bumbaco, A. E., Kellam, N., Pawley, A. L., & Phillips, C. M. L. (2017). Qualitative Research Quality: A Collaborative Inquiry Across Multiple Methodological Perspectives. *Journal of Engineering Education*, *106*(3), 398–430. https://doi.org/10.1002/jee.20170

Walther, J., Sochacka, N. W., & Kellam, N. N. (2013). Quality in interpretive engineering education research: Reflections on an example study. *Journal of Engineering Education*, *102*(4), 626–659. https://doi.org/10.1002/jee.20029

Whitson, J., Eaket, C., Greenspan, B., Tran, M. Q., & King, N. (2008). Neo-immersion: Awareness and engagement in gameplay. *ACM Future Play 2008 International Academic Conference on the Future of Game Design and Technology, Future Play: Research, Play, Share*, 220–223. https://doi.org/10.1145/1496984.1497028

Willey, R. J. (1999). SACHE Case Histories and Training Modules. *Process Safety Progress*, *18*(4), 195–200. https://doi.org/10.1002/prs.680180405

Willey, R. J., Carter, T., Price, J., & Zhang, B. (2020). Instruction of hazard analysis of methods for chemical process safety at the university level. *Journal of Loss Prevention in the Process Industries*, *63*(November 2018), 1–9. https://doi.org/10.1016/j.jlp.2019.103961

Woodzicka, J. A., & LaFrance, M. (2001). Real versus imagined gender harassment. *Journal of Social Issues*, *57*(1), 15–30. https://doi.org/10.1111/0022-4537.00199

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, *82*(1), 61–89. https://doi.org/10.3102/0034654312436980

Zhu, Q., Zoltowski, C. B., Feister, M. K., Buzzanell, P. M., Oakes, W. C., & Mead, A. D. (2014). The development of an instrument for assessing individual ethical decision-making in project-based design teams: Integrating quantitative and qualitative methods. *ASEE Annual Conference and Exposition, Conference Proceedings*.

Zuparic, M., Khuu, D., & Zach, T. (2021). Information theory and player archetype choice in Hearthstone. *Information Sciences*, *559*, 236–250. https://doi.org/10.1016/j.ins.2021.01.066