2017

Improving Planning: Quantitative Evaluation of the Premortem Technique in Field and Laboratory Settings

Madeline Peabody
Michigan Technological University, mjpeabod@mtu.edu

IMPROVING PLANNING: QUANTITATIVE EVALUATION OF THE PREMORTEM TECHNIQUE IN FIELD AND LABORATORY SETTINGS

By

Madeline J. Peabody

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2017

This thesis has been approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE in Applied Cognitive Science and Human Factors.

Department of Cognitive and Learning Sciences

Thesis Advisor: *Dr. Elizabeth Veinott*

Committee Member: *Dr. Shane Mueller*

Committee Member: *LTC Adam Melnitsky*

Department Chair: *Dr. Susan Amato-Henderson*

Table of Contents

List of Figures

List of Tables

Preface

The following thesis is intended for publication. The data collection and analyses are my original work under the guidance of Dr. Elizabeth Veinott. I wrote this manuscript under the review of Dr. Elizabeth Veinott.

Acknowledgements

This research and publication would not have been possible without the guidance and mentorship of my advisor, Dr. Elizabeth Veinott. Beth: thank you for the countless hours spent going over my data, proofreading my writing, and practicing for talks and presentations. Without your continuous support and patience, I would not be where I am today.

I am also very grateful for my committee who spent time reading this thesis and making comments and suggestions. LTC Adam Melnitsky and Dr. Shane Mueller: you each played a substantial role in my undergraduate success; I am thankful to have you on my committee and for all your guidance over the years.

I also want to thank my friends and family for their overwhelming encouragement throughout my academic career. Mom, Dad, Erika, Phil, and all my sisters: words cannot describe how happy and fortunate I feel to have you in my life; your support emboldens me and inspires me to succeed. Thank you for always being there for me.

Abstract

Planning can be difficult and developing techniques for evaluating plans has been limited. This thesis compares different plan evaluation techniques in a series of experiments. The main techniques discussed are the Premortem Method and Worst-Case Scenario Method. The Premortem plan evaluation method can help people reduce overconfidence and generate more reasons a plan might not succeed. Only one experiment has validated this technique; therefore, one goal of the present series of experiments is to qualitatively and quantitatively examine the effectiveness of the Premortem Method in several different planning situations. This research evaluates the extension of the Premortem to shorter planning time periods, evaluates the effectiveness with team generated and executed plans, and compares the use of this technique among individuals and teams. In Experiment 1, 52 Army Cadets operating in teams completed six time-constrained field exercises that required planning, half using the Premortem and half using a standard Military plan evaluation process. Compared to a control condition, when teams used the Premortem they had fewer fouls and less fixation with no change in planning and execution time. In Experiment 2, 72 individual participants from university organizations used the Premortem Method or Worst-Case Scenario Method to evaluate their group's plan for an engineering task. Results from Experiment 2 indicated that there was no statistically significant difference in the number of reasons and solutions generated between methods. However, the two methods had significantly different distributions of reasons and solutions across categories, indicating that the methods were prompting participants to approach the plan differently. To further examine the relative

effectiveness of these two plan evaluation methods, and the influence of group dynamics, Experiment 3 compared the efficacy of the Premortem and Worst-Case Scenario Method amongst groups and individuals in face-to-face settings with a complex and unfamiliar plan. Eighty-two participants generated more reasons with the Premortem Method than the Worst-Case Scenario Method, and groups generated more solutions than individuals did. Overall, the participants in groups using the Premortem Method produced more reasons and solutions than participants using the Worst-Case Scenario Method and individual participants using the Premortem Method. The distribution of reasons was also significantly different across categories, indicating that the underlying mechanisms are changing how participants view the problem and generate reasons. These studies extend prior work by validating that the Premortem is effective in short planning horizons, demonstrating that it works for individuals and teams, and clarifying potential boundary conditions. This research advocates several directions for future research, and suggests possibility of future implementation as a virtual tool or application.

Chapter 1: Introduction

Planning is difficult for many reasons. Problems can emerge, be difficult to understand, and hard to solve because they are poorly defined (Rittel & Webber, 1973). Military decision making frequently occurs in these uncertain environments; information might be missing, intelligence reports might be contradictory, and time might be running out. Leaders must consider several different factors and seek to fully understand the complexity of problem situations when making command decisions (Cojocar, 2011). The U.S. Army formally utilizes two tactical planning processes: The Military Decision Making Process (MDMP) and Troop Leading Procedures (TLP). MDMP is used as a framework for guiding leadership decisions and planning at higher echelons where Commanders have staff support; TLP is an abbreviated version of MDMP used at lower echelons such as company or squad (Army Planning and Orders Production, 2005). There are several steps to MDMP, but the process is designed to be a flexible plan for adaptive, intuitive leaders. Military leaders must make hunches about the world and imagine causes and effects; and steps can be added and subtracted to the process as needed. Alongside their staff and other key leaders, Commanders can utilize these platforms to gain a better understanding of the operational climate through discussion and wargaming (Perez, 2011). While MDMP and TLPs are useful, flexible frameworks for planning, they can be time-consuming and difficult to understand. The Army is missing a lightweight planning tool to incorporate in uncertain, time-constrained environments. One goal of this thesis is to evaluate the efficacy of a modified Premortem (Klein, 2007) in a naturalistic field setting to bridge this gap.

The Army is not the only organization with a need for a lightweight, easy-to-use planning tool. Organizations use myriad tools, techniques, and approaches to planning and decision-making. Many strategies assess goals and plans at the organizational level, but the successes of these approaches remain largely anecdotal (Meissner & Wulf, 2015). Many plan evaluation techniques lack quantitative validation; it is difficult to retrospectively examine their effectiveness due to the complexities of decision making and planning in complex environments (Wright, Bradfield, & Cairns, 2013). The current series of experiments will contribute quantitative data regarding the evaluation of both real and hypothesized plans with groups and individuals.

By utilizing a failure frame and reducing uncertainty, the Premortem can help leaders and members of a team reduce overconfidence and identify holes in their plan prior to execution. The Premortem technique is a validated five-step method based on key findings in problem solving and decision making used to evaluate existing plans (Klein, 2007). The key to this approach is to evaluate the plan as if it has completely failed, and looks for reasons why it failed and solutions to mitigate those potential failures. The five key steps in the Premortem are outlined in Figure 1.1.

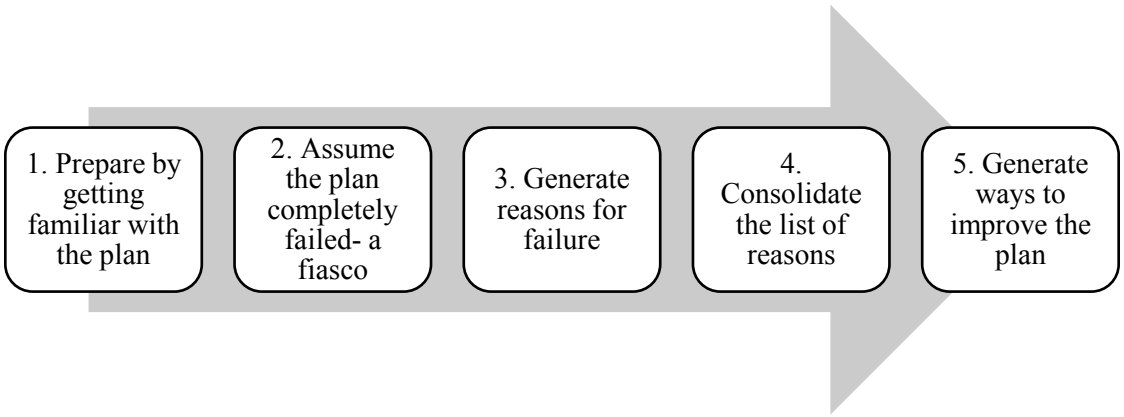

Figure 1.1. The steps in the Premortem plan evaluation process. Adapted from Klein, Performing a Project Premortem, 2007.

Chapter 2: Literature Review

Military Decision Making

Classical analytical models of decision making often encourage an organized approach of weighing options, evaluating them, and choosing the best one (Saaty, 2008). However, these models operate under the assumption that the user understands the problem and state of the field. Military environments can be ambiguous and uncertain, and thus are more compatible with Naturalistic Decision Making strategies: goals can be ill-defined, conditions can rapidly change, and time can be critical (Klein, 2008). MDMP in an analytical model developed to be appropriate in tactical operations, and can help commanders counteract psychological traps and biases (Marr, 2000). However, military professionals are dissatisfied with the process; it is lengthy and subsequently might not be utilized properly. In theory and in research, MDMP is an effective tool that enhances combat-action decision making and flexible planning strategies (Marr, 2000). Nonetheless, commanders would benefit from utilizing a more lightweight method in developing their plans.

Overconfidence in Planning

People tend to be overconfident in their judgments, beliefs, and predictions and do not naturally question how they arrived at a decision or how other alternatives stack up (Fischhoff, Slovic, & Lichtenstein, 1977). The Premortem can help reduce overconfidence in groups and individuals during the planning process. There is both a value and danger to overconfidence; if people are not confident, they are less likely to

take on risky and potentially rewarding projects. Yet too much confidence can cause humans to overestimate the potential of plans, people, and projects (Russo & Shoemaker, 1992). People tend to believe they will complete tasks sooner than they actually do, and are overconfident in their assessment of their plan completion times (Buehler, Griffin, & Ross, 1994). This planning fallacy can pose a problem for individuals and organizations operating on deadlines. The Premortem can provide a solution to the planning fallacy by reducing overconfidence.

Overconfidence intensifies in group dynamics. While groups tend to be more accurate than individuals, they are more susceptible to overconfidence in their incorrect answers (Puncochar & Fox, 2004). This poses a problem for organizations and project teams where group members rely on each other for ideas, critique, and feedback of the overall plan. When people are relying on heuristics to make decisions, confidence spikes; there are several methods for debiasing that can reduce overconfidence (Brenner & Koehler, 1996). Premortem reduces overconfidence by utilizing a failure frame and certainty of outcome (Veinott, Klein, & Wiggins, 2010).

The Effects of Frames

How a problem is framed affects our evaluations and decisions. Tversky and Kahneman's Prospect Theory describes that people treat losses as being much worse than they treat gains of the same magnitude (1981). When people are challenged to imagine or explain why a hypothesis is true, they are more confident due to the conditional reference frame, and this frame effects their evaluative processes. Framing influences

how people interpret relevant evidence and how they seek out and find information (Koehler, 1991). Changing their frame from optimism to pessimism can have profound effects on confidence and how they approach the problem. Different frames elicit different emotions, and the emotional system is involved in decision-making (De Martino, Kumaran, Seymour, & Dolan, 2006). If a decision or plan is framed in terms of failure, people will look at the decision more subjectively and with less confidence.

Evidence for Debiasing Confidence

People have biases to favor positive evidence and to disregard evidence that is inconsistent with their answers or beliefs. When asked to generate reasons for their answers or beliefs, confidence calibration improves (Koriat, Lichtenstein, & Fischhoff, 1980). In a study with neuropsychologists and diagnosis of a hypothetical patient, researchers found that asking the professionals to list reasons for their selected diagnosis can debias and subsequently reduce overconfidence (Arkes, Guilmette, Faust, & Hart, 1988). Alternative possibilities seem plausible after imagining potential complications that might occur in a given situation. This also leads to a lower manifestation of hindsight bias.

Hindsight and Uncertainty

Outcome knowledge is refered to as hindsight; after gaining hindsight, people tend to falsely believe that they could have predicted the outcome before it occurred. This is known as the hindsight bias (Hawkins & Hastie, 1990). The prospective hindsight mechanism can be used in decision making and plan evaluation to counteract this effect

(Mitchell, Russo, & Pennington, 1989). This mechanism brings future events into the present tense; in doing so, people are able to understand potential problems and implications better. Mitchell and colleagues found by imagining a future event has already occurred, and examining why it happened, people will generate 30% more reasons an event may occur (1989). Frame and uncertainty can be manipulated to help people see problems before they arise.

Decision Making: Idea Generation and Plan Evaluation

Brainstorming. Problem-solving procedures often include idea generation through brainstorming exercises. The most effective problem-solving strategies involve individual brainstorming procedures; when people analyze problems on their own and then discuss their ideas with a group, they generate more reasons than groups brainstorming together (Bouchard, 1969). Brainstorming groups are most effective with five members; if the group is larger, not all group members have a chance to speak and discuss their ideas (Bouchard & Hare, 1970). Brainstorming encourages wild ideas and does not allow criticism or adverse judgement; ideas can be combined and groups can reach solutions more efficiently (Bouchard, Drauden, & Barsaloux, 1974). The Premortem operates under the same principles; people generate reasons for failure on their own, discuss the reasons with the group, generate solutions on their own, and discuss the solutions with the group. Through individual brainstorming, people generate more unique ideas (Bouchard, 1969). Reasons can be combined and expanded through discussion, and group members can generate solutions for others' reasons.

The Scenario Planning Method. Scenario planning has long been used to evaluate plans (Chermack, 2004). Scenario planning methods can be used as a framework for large organizations to develop a greater understanding of their overall goals, potential outcomes, and strategies to reach desired end states. However, scenario planning methods, and enhancements to these methods (including the backwards logic, intuitive logics, and antifragile methods) fail to control for both framing affects and uncertainty reduction (Wright, Bradfield, & Cairns, 2013). Many scenario methods operate on the basis that the future is unknown; to make predictions, one must imagine different scenarios, both good and bad, that might occur (Wright & Goodwin, 2009). Constructing these forced scenarios helps stakeholders examine all potential positive outcomes alongside all possible negative outcomes. At a macro-level, this approach aims to stretch people's thinking individually and collectively by building multiple scenarios and presenting these scenarios as possibilities rather than predictions (Shoemaker, 1993). Users imagine their best- and worst-case scenarios, and develop a story of how those scenarios might unfold. These methods use backwards logic and causal reasoning.

The Strategy Scenario Approach. Meissner and Wulf (2015) suggested a new approach to enhance scenario planning methods; the strategy scenario approach has four steps to analyze contingencies and develop strategies. At a macro-level, this approach aims to incorporate decision-making processes with scenario-based reasoning and planning. An important step in this process is to develop flexible strategies; this is done by performing a Premortem analysis (Meissner & Wulf, 2015).

The Premortem. The Premortem adapts research in problem solving and decision making in order to reduce uncertainty and bring hindsight into the present tense. This plan evaluation technique utilizes the prospective hindsight mechanism (Mitchell, Russo, & Pennington, 1989) and a failure frame (Tversky & Kahneman, 1981). When people approach a scenario with a failure mindset and uncertainty reduction, they generate more reasons the plan will fail, and say things they ordinarily would not suggest (Klein, 2007).

The Premortem has the potential to work in many different environments, but has not yet been thoroughly experimentally evaluated. So far only one study has attempted to validate the strategy; research conducted by Veinott, Klein, & Wiggins (2010) found that the Premortem shows a greater reduction in confidence than the Pro/Con, Con Only, and Critique plan evaluation methods. The Premortem has potential to be quantitatively studied, and can be practically integrated in an organizational-level strategy scenario method as a tool to reduce uncertainty and improve plans (Meissner & Wulf, 2015). Along with quantitative confidence reduction, the Premortem Method has anecdotal evidentiary support, but have not been tested systematically with teams who are generating their own plans.

Present Research

The present research is comprised of three experiments aiming to examine the Premortem in several systematic ways, each time leveraging aspects of real-world planning. Specifically, it is hypothesized that the Premortem will be superior to other

plan evaluation methods in reducing overconfidence, generating more plan problems and ways to improve or strengthen plans, and achieving better outcomes.

In Experiment 1, the Premortem was qualitatively and quantitatively tested in a time-constrained environment (approximately 25 minutes) with emerging problems. This was done by implementing the technique with Reserve Officer Training Corps (ROTC) Cadets at Michigan Technological University (MTU) during their field leadership reactionary course (FLRC). Validation in a field environment and shorter time frame suggests that event a brief and lightweight Premortem can be effective. Next, for Experiment 2, the Premortem was directly compared to an artifact of the Scenario Planning Method: construction of a "Worst-Case Scenario." This experiment was conducted with pre-existing groups using their own real plans; groups participating in MTU's Winter Carnival month-long statue building event were asked to evaluate their organization's snow statue plans. Distinction between the Worst-Case Scenario and the Premortem methods differentiates the two methods; number of problems discussed and solutions generated determines the effectiveness of the Premortem's utilization of prospective hindsight. Lastly, in Experiment 3 the Premortem Method and Worst-Case Scenario Method were compared between face-to-face groups and individuals to further explore the distinction between the two methods and the effectiveness with groups versus individuals. Examining the Premortem technique in these different settings will help establish boundary conditions for its effectiveness, and provide insight for future implementations.

Chapter 3: Experiment 1

The Premortem has previously been implemented in formal, structured settings. No research has examined the method's potential as an effective, lightweight planning tool in a dynamic, time-critical environment. Experiment 1 evaluated whether a lighter version of the Premortem was more effective than the current Military field planning practice, and whether the results replicated previous findings (Veinott, Klein, & Wiggins, 2010). There are several important distinguishing factors between Experiment 1 and the research conducted by Veinott and colleagues; in Experiment 1, participants used their own plans, teams operated in a shorter timeframe, each team generated and evaluated multiple plans using each technique, and plans were executed. Einstellung, a habituation to repeatedly used procedures, was hypothesized to occur on the obstacles (Luchins, 1942). This mechanization in problem solving is referred to as problem fixation. In addition to change in confidence, Experiment 1 examined the effects of a lighter version of the Premortem on problem fixation and number of fouls during the execution of the plan.

"The material contained in this chapter is in preparation for submission to a journal".

Hypotheses

Compared to the Baseline planning technique, teams using the Premortem Method will:

H1. Have a greater change in confidence.

H2. Have less problem fixation.

H3. Make fewer fouls during execution.

Methods

Participants. Fifty-eight ROTC Cadets were recruited to participate in this study. The majority (81%) were male, and the average age was 19.6 years (*SD*=1.06). Participation was part of the Cadets' semi-annual field training exercise. They were organized into six pre-determined teams comprised of seven or eight members, with experience levels ranging from a few weeks to three years and evenly distributed across teams.

Observers. Twelve fourth-year Cadets served as observers of the team performance and facilitators of the plan evaluation; six remained stationary at each obstacle, and six were assigned to teams. For each obstacle, both observers rated teams' fixation. Their ratings were significantly correlated, $r(34)=.595$, $p<.01$. They also each rated outcome and teamwork; only obstacle observers recorded the number of fouls. The six teams were labeled by color: Black, Red, Green, Blue, Yellow, and Orange (Table 3.1).

Table 3.1

Counterbalancing design for Experiment 1. Shading indicates Premortem Method.

Team/Rotation	1	2	3	4	5	6
Black	Quicksand	Elevation	Medivac	Demolition	Recon	Wall Banger
Red	Wall Banger	Quicksand	Elevation	Medivac	Demolition	Recon
Green	Recon	Wall Banger	Quicksand	Elevation	Medivac	Demolition
Blue	Demolition	Recon	Wall Banger	Quicksand	Elevation	Medivac
Yellow	Medivac	Demolition	Recon	Wall Banger	Quicksand	Elevation
Orange	Elevation	Medivac	Demolition	Recon	Wall Banger	Quicksand

Design. The experiment was a modified Latin Squares design. Plan evaluation method was counterbalanced across plans, with each team attempting three obstacles with each technique. The independent variables were method (Baseline, Premortem) and obstacle (Quicksand, Elevation, Medivac, Demolition, Recon, Wall Banger). The dependent measures, defined in Table 3.2, were confidence, fouls during execution, fixation, teamwork, and outcome. Time was also examined (planning, execution, total time).

Table 3.2

Dependent Measures of Team Performance

Confidence	How confident are you that your group's plan will be successful? *Rate this on a scale of 0-100%.*
Fixation	Do they keep trying the same maneuver even though it keeps failing? How many times do they try it before moving on to a different technique? How long were they stuck at this phase? *Rate this using a scale of 1-6 (1 minimal fixation, 6 high fixation)*
Teamwork	How well are they working together as a team during the planning and execution stages? *Rate this using a scale of 1-6 (1 poor teamwork, 6 excellent teamwork)*
Outcome	Did they complete the outcome in the allotted time? Did they complete it quickly and with few errors? *Rate this using a scale of 1-6 (1 poor outcome, 6 excellent outcome)*
Fouls	Obstacle observers only: count the number of pre-determined fouls (outlined in handbook) teams make on each obstacle. *Touch black areas, drop equipment, etc.*

Materials. This experiment was conducted outdoors on an obstacle course at Ft. McCoy, Wisconsin. There were six obstacles used in this exercise: Quicksand, Elevation, Medivac, Demolition, Recon, and Wall Banger. Team members and observers were given color-coded booklets and packets that corresponded to their team name. All data was recorded in these booklets and packets, and later transcribed by two researchers.

Procedure. The basic procedure is outlined in Figure 3.1. A different leader was selected by the team observer for each obstacle. The leader received the mission from the obstacle observer and made a plan. The leader briefed the plan to the team. All team members recorded the leader's plan. In the Premortem condition, each team member

recorded reasons the plan would fail and solutions for those reasons. In the Baseline condition, which constituted the standard Military practice for this type of field planning situation, each team member recorded any questions they had about the plan. In both conditions, participants rated their confidence two times rather than three times as had been done previously (Veinott, Klein, & Wiggins, 2010), due to the short timeframe. Confidence was initially rated after the leader briefed the plan, and then rated again after generating reasons/solutions or asking questions about the plan (Figure 3.1). Black, Green, and Yellow teams began using the Premortem Method; the other three did not deviate from standard procedure (Baseline). After completing three obstacles, the teams changed methods and the Red, Blue, and Orange teams negotiated obstacles using the Premortem Method.

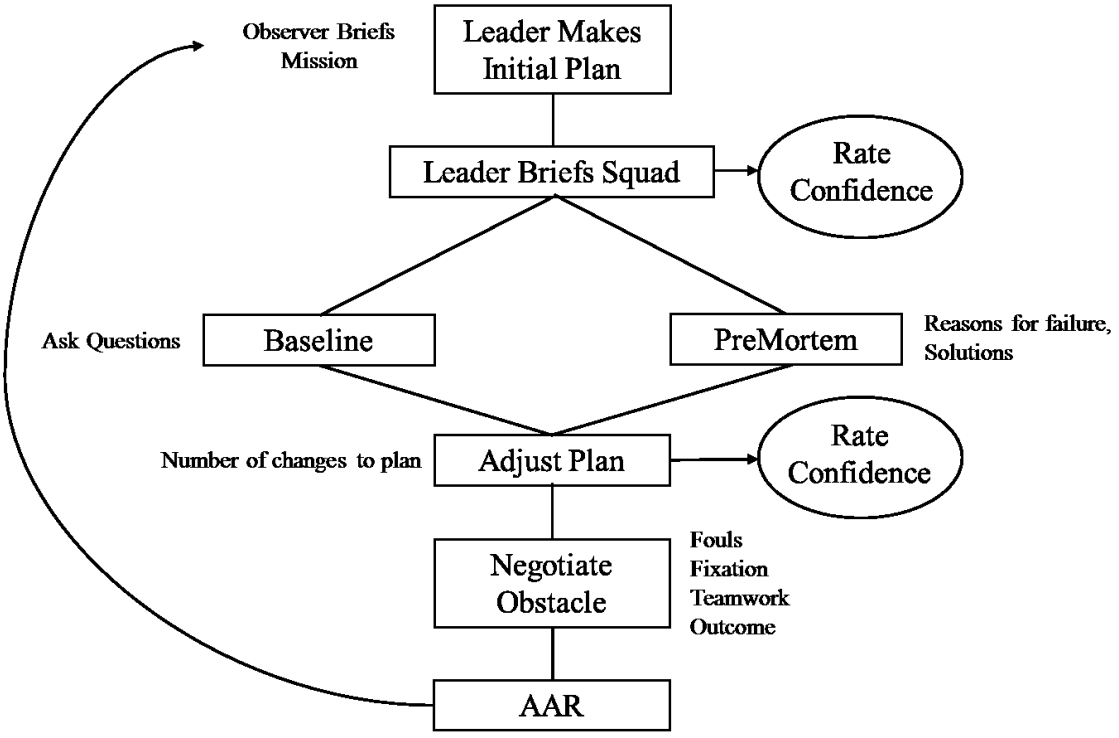

Figure 3.1. Procedure for Experiment 1.

Some of the dependent measures were captured by a team observer. All observers were given a packet in which they recorded the same information: the leader's plan, their own first confidence rating, reasons/solutions or questions discussed, their own second confidence rating, changes to the plan, and fouls, fixation, teamwork, and outcome during execution. Observers also detailed the time the leader received the mission, the time the leader briefed the mission to the team, time execution began, and time execution ended. During execution, observers recorded the number of fouls and gave subjective ratings on problem fixation, teamwork, and outcome. They also took notes during execution, describing how the team negotiated the obstacle.

Coding Scheme. Reasons and solutions were each divided into four corresponding categories (Table 3.3). The categorization of reasons and solutions was conducted for descriptive purposes. The categories were established after reading and discussing the types of reasons and solutions, and two independent raters assigned each reason and solution to a category. Two hundred and seventy reasons and solutions were categorized. The raters established broad criteria for the categories; if there was disagreement or confusion between the raters, the specific reasons and categories were discussed and the category criteria were updated. Cohen's Kappa, which controls for chance, was used to determine inter-rater reliability. For this coding scheme, the raters had high agreement (K=.817) (McHugh, 2012). The categories are represented in Table 3.3 with examples from the data; further analyses of categories are discussed in the Results section.

Table 3.3

Experiment 1 categories of reasons and solutions with descriptions.

Reasons	Solutions
Order • Person can't get over in beginning/end • Can't get the first person across • Not being able to boost last person up on post	**Order** • Person sitting on board will assist final person • Send biggest guy over first • Do the reverse of the first method
Predetermined fouls • Knock post with legs • Pole will touch fence • Someone touches the gray area	**Situation awareness** • Be careful • Be mindful of post • Caution and communication
Equipment • Overlapping boards could fall • Wagon won't fit • Ammo can is heavy, could fall	**Equipment** • Assembly line for equipment • Tie boards together with rope • Use pulley system to move ammo
Physical ability • Someone will fall off the inclined boards • Not everyone can climb a rope • Might have trouble holding and people might slip	**Jobs and tasks** • Cadets support each other on the way up • Stage people along route • Lift short people up • Have a person on rope guard

Results

Data were analyzed using univariate and repeated measures ANOVA. Planned and post-hoc t-tests were conducted to evaluate specific hypotheses and follow up on statistically significant interactions. All statistical tests used an alpha level of .05, and

exact p-values are reported. P-values were corrected using Tukey's HSD, and Greenhouse-Geisser corrections were used when Mauchly's test of sphericity indicated assumptions were violated (the F-statistic is biased and therefore invalid; there is an increased chance of a false positive, so degrees of freedom must be corrected). Results first examine specific hypotheses for each experiment, then delve into further analyses. Results supported H2 and H3; however, H1 was not supported (Table 3.4).

Table 3.4

Experiment 1 hypotheses and results.

H1: Confidence, *not supported*	Individuals did not have a greater change in confidence when using the Premortem Method.
H2: Problem Fixation, *supported*	Teams had less problem fixation when using the Premortem Method.
H3: Fouls, *supported*	Teams had fewer fouls during execution when using the Premortem Method.

H1: Confidence. A 2x2 repeated measures ANOVA was conducted to compare the effect of method (Baseline, Premortem) and rating time [time 1 (baseline), time 2 (after plan evaluation manipulation)] on confidence ratings (0-100%). There was a significant main effect of rating time on confidence, $F(1,243)=133$, $p<.001$, Wilk's $\Lambda = .646$, $\eta_p^2 = .354$; indicating that team members' confidence increased between ratings 1 and 2. However, the main effect of method on confidence was not statistically significant, $F(1,243)=.376$, $p=.540$, Wilk's $\Lambda = .998$, $\eta_p^2 = .002$; in both cases, confidence

was higher in the second rating, indicating that Premortem failed to change individual confidence ratings statistically more than the Baseline method did (Figure 3.2).

An independent samples t-test was used to examine change in confidence between the methods. The change in confidence between times 1 and 2 for participants using the Premortem Method (M=7.56, SD=9.93) was not statistically different from the Baseline (M=7.30, SD=9.25), $t(311)$=.237, p=.813. These mean confidence ratings were across all participants for all obstacles.

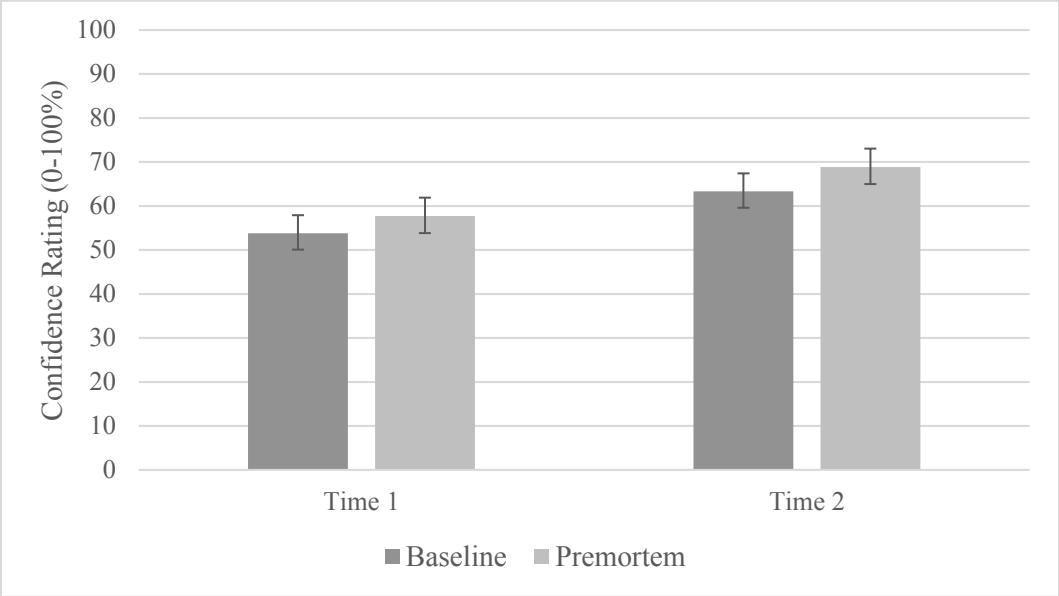

Figure 3.2. Mean and standard error of confidence rating by plan evaluation method and time.

H2: Fixation. For each obstacle, both observers rated teams' fixation. Across teams, these observer ratings were statistically correlated, $r(34)$=.595, p<.01. A 2x6 factorial ANOVA examining fixation by method (Baseline, Premortem) and obstacle (Quicksand, Elevation, Medivac, Demolition, Recon, Wall Banger) resulted in main

effects of method, $F(1,69)=4.44$, $p=.040$, $\eta_p^2= .071$; and obstacle, $F(5,69)=4.30$, $p=.002$, $\eta_p^2= .270$. This indicates that teams were less fixated when using the Premortem Method, and fixation was significantly varied amongst obstacles. The interaction was not significant, $F(5,69)=1.18$, $p=.329$; $\eta_p^2= .093$ (Figure 3.3). Given the variability across the obstacles, they were separated into well-defined and ill-defined problem types for further analyses.

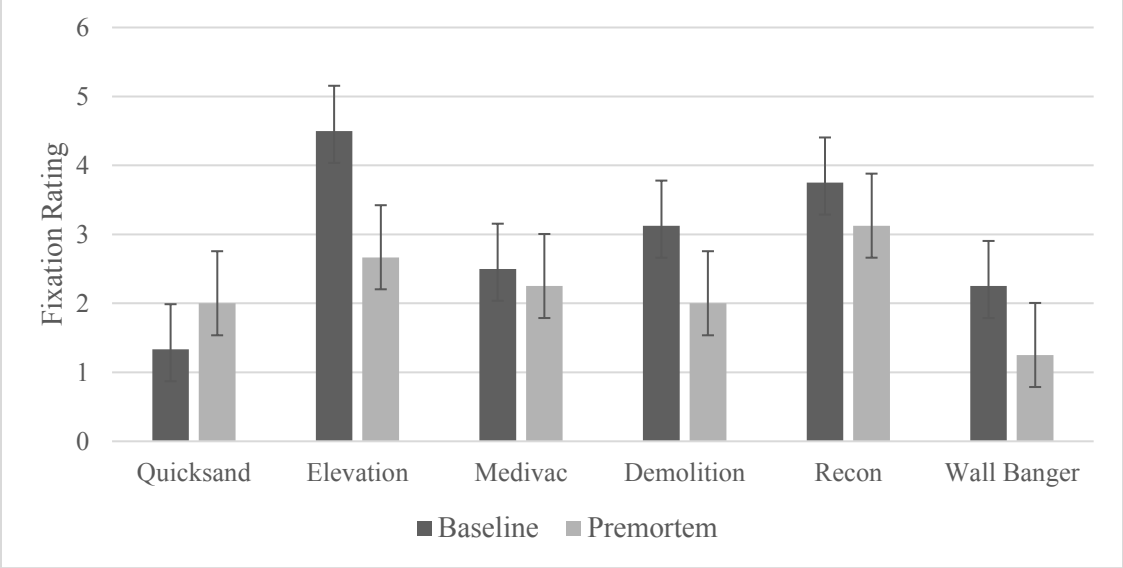

Figure 3.3. Fixation rating by plan evaluation method and obstacle.

Fixation by Problem Type. Well-defined problems had clear strategies and methods; most teams approached the obstacle in similar ways and had higher outcome scores. Outcomes were rated by observers on a scale of one (poor outcome) to six (excellent outcome). Quicksand had the best outcomes ($M=5.55$, $SD=.688$); followed by Medivac ($M=5.50$, $SD=.674$), Demolition ($M=4.82$, $SD=.982$), and Wall Banger ($M=4.67$, $SD=1.23$). Quicksand, Medivac, Demolition, and Wall Banger were classified as well-

defined problems. Ill-defined problems were more difficult to solve and teams attempted several different techniques. Elevation (M=2.08, SD=1.24) and Recon (M=3.08, SD=2.02) had worse outcomes; Elevation and Recon were classified as ill-defined problems. An independent samples t-test showed obstacles categorized as ill-defined problems (M=2.58, SD=1.72) had significantly lower outcomes than well-defined problems (M=5.13, SD=.980), $t(68)=7.92$, $p<.001$.

A 2x2 factorial ANOVA examining fixation by method (Premortem, Baseline) and problem type (well-defined, ill-defined) found main effects of method on fixation, $F(1,69)=6.12$, $p=.016$, $\eta_p^2=.085$, indicating that teams were less fixated when using the Premortem Method. There was also a main effect of problem type on fixation, $F(1,69)=17.7$, $p<.001$, $\eta_p^2=.211$, indicating that teams had significantly higher fixation when negotiating ill-defined problems. The interaction was not significant, $F(1,69)=1.69$, $p=.199$, $\eta_p^2=.025$.

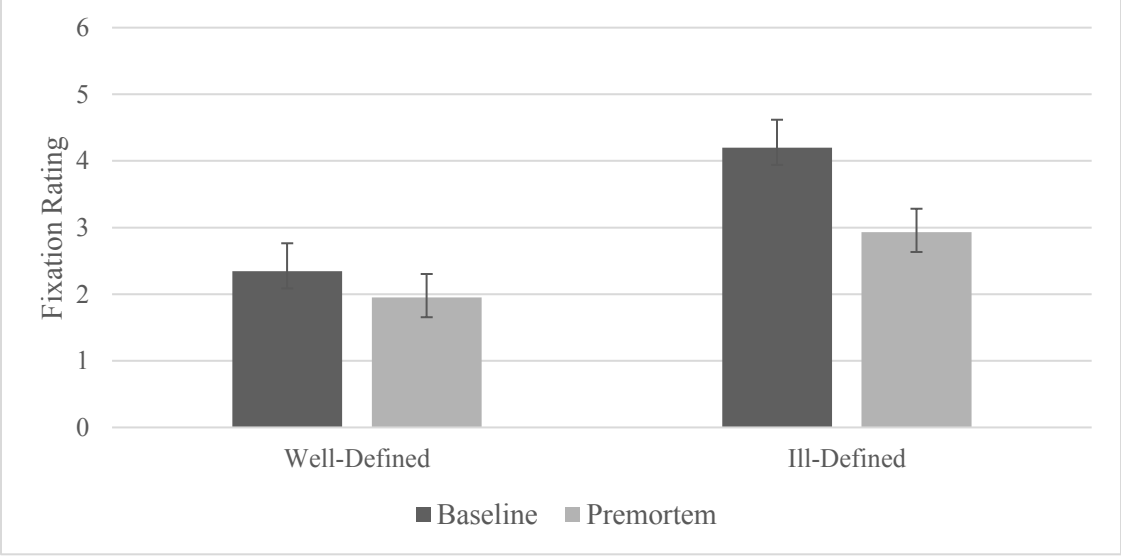

Figure 3.4. Fixation rating by plan evaluation method and problem type.

H3: Fouls. A 2x6 factorial ANOVA examining number of fouls by method (Baseline, Premortem) and obstacle (Quicksand, Elevation, Medivac, Demolition, Recon, Wall Banger) found a main effect of method, $F(1,35)=5.55$, $p=.027$, $\eta_p^2=.188$, indicating that teams had fewer fouls in when using the Premortem Method. There was also a main effect of obstacle, $F(5,35)=15.5$, $p<.001$ $\eta_p^2=.764$, indicating that some obstacles consistently had far more fouls than others. The interaction was not significant, $F(5,35)=1.66$, $p=.184$, $\eta_p^2= .257$.

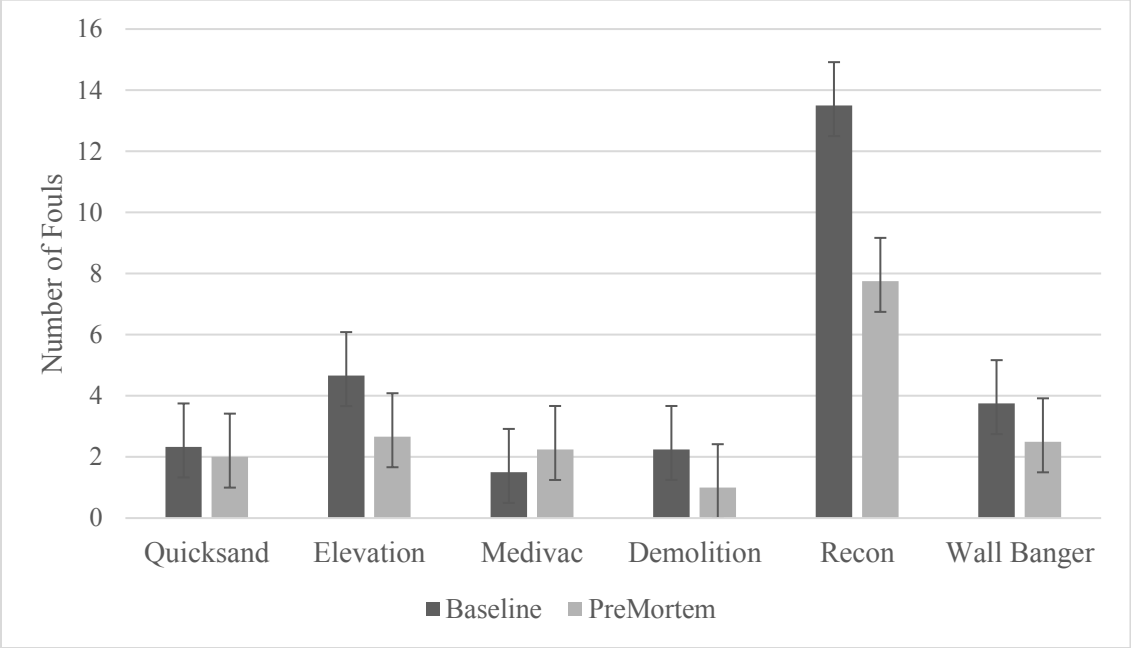

Figure 3.5. Number of fouls by plan evaluation method and obstacle.

A 2x2 ANOVA examining number of fouls by method (Baseline, Premortem) and problem type (ill-defined, well-defined) revealed a main effect of problem type $F(1,35)=21.4$, $p<.001$, $\eta_p^2=.401$, indicating that ill-defined problems had far more fouls than well-defined problems. There was no significant difference in number of fouls by

method, $F(1,35)=2.69$, $p=.111$, $\eta_p^2=.077$, indicating that regardless of which plan evaluation method they used, teams had about the same number of fouls on the same problem types. The interaction of method and problem type was not significant, $F(1,35)=1.03$, $p=.317$, $\eta_p^2=.031$. These results differ from the number of fouls by obstacle because obstacles were not evenly distributed across problem types; there were two ill-defined problems and four well-defined problems. With a larger sample size, this effect might reach significance.

Number of Reasons and Solutions. Participants generated 288 reasons for plan failure when evaluating their plans using the Premortem Method. The number of reasons varied by obstacle and zero to five, with a mean of about 2 reasons per person ($M=2.12$, $SD=1.14$). Of the reasons for failure generated, the majority (43.1%) were related to equipment, followed by physical ability (24.3%), predetermined fouls (23.6%), and order (9%).

There were 234 solutions generated, ranging from zero to seven per individual per obstacle ($M=1.78$, $SD=1.25$). Of the solutions generated, the majority (51.7%) were related to equipment, followed by assigning jobs and tasks (26.1%), situation awareness (19.7%), and order (2.6%).

Time. Obstacle observers recorded how long teams spent in their planning and execution phases. Multiple independent sample t-tests were conducted to examine time. In the planning phase, teams spent fewer minutes in the Baseline ($M=8.77$, $SD=2.28$) than Premortem ($M=10.5$, $SD=3.20$); this was a marginally statistically significant difference,

t(34)=-1.86, *p*=.072. During execution, there was no statistical difference in time between the Baseline (*M*=16.4, *SD*=4.45) and the Premortem (*M*=16.2, *SD*=2.90), *t*(34)=.177, *p*=.861. Overall, teams spent the same amount of time in the Baseline (*M*=25.2, *SD*=4.85) and the Premortem (*M*=26.7, *SD*=4.03), *t*(34)=-1.01, *p*=.322. While the Premortem added time to the planning process, it did not add any time to the execution and overall time that teams spent negotiating the obstacle.

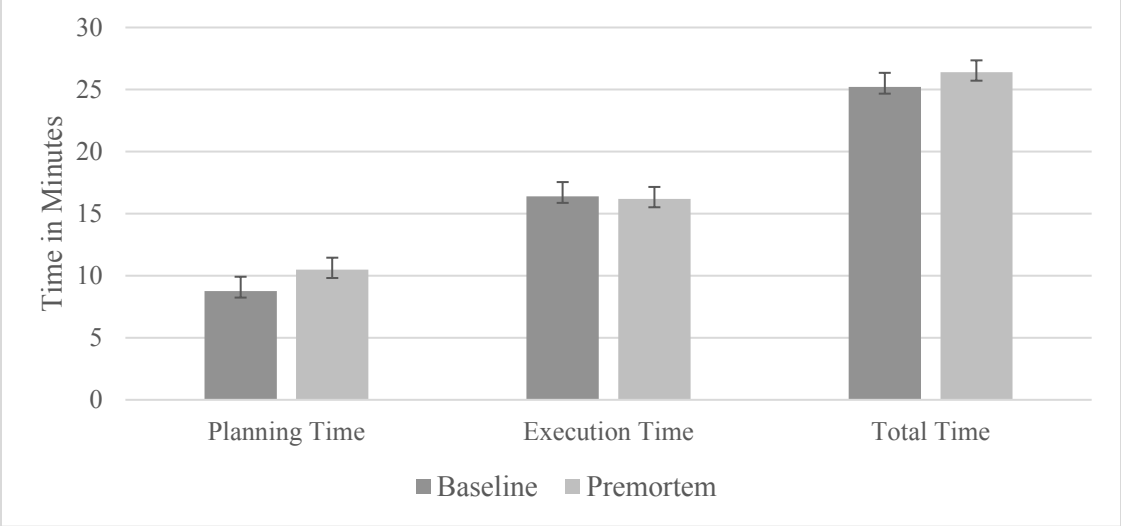

Figure 3.6. Time by planning method.

Post-Survey. Participants were asked to fill out a post-questionnaire at the end of the day (Appendix A). A Likert scale from 1 (not at all) to 9 (a lot) was pictured, and participants filled in the survey. Fifty-seven surveys were collected. After the last round, participants found the Premortem to be easy to use in this type of planning environment (*M*=6.89, *SD*=1.48). They also thought their squads generated good reasons for plan failure (*M*=6.33, *SD*=1.66). The Premortem marginally changed their planning process (*M*=5.54, *SD*=1.97), and somewhat improved solutions during the planning process

(*M*=5.82, *SD*=1.61). Relative to the baseline practice, they believed the Premortem only somewhat affected their team's ability to recover from a bad idea (*M*=5.61, *SD*=2.23). While participants did not fully believe the Premortem was helping them, they did find it easy to use. Nonetheless, the method was working; teams had fewer fouls and less fixation when using the Premortem Method.

Discussion

Overall, several of the current hypotheses were supported. While there was no significant change in confidence by method, participants had significantly lower problem fixation and fewer fouls when using the Premortem Method. Fixation was higher for ill-defined problems; when problems are difficult to solve, participants often tried the same method repeatedly even though it was not working. It is likely that they could not generate better solutions for the issues they were encountering during execution. On the other hand, fixation was significantly lower in the well-defined problems because they ran into fewer issues during execution. Their method worked, so they were not trying and failing the same method repeatedly.

Participants' confidence did not change more when using the Premortem Method, as was hypothesized. This could be attributed to a misunderstanding of the use of the confidence scale (Adams, 1957), group dynamics and development throughout the exercise, or time constraints of the ratings coming within a few minutes of each other.

The Premortem Method was more effective than the Baseline Method and did not extend the total time for planning and execution. While the Premortem Method added a

small amount of time during planning, it reduced fixation and produced fewer fouls. There was no significant difference in overall time between the Premortem and Baseline. Decreased fixation and fouls when using the Premortem Method indicates that the Premortem helps teams develop plans that are more resilient and flexible. In many cases, after going through the Premortem, leaders did not make significant changes to their overall plan. However, during the execution phase, groups that discussed reasons their plan might fail and solutions for those problems were less likely to keep attempting an unsuccessful technique on the obstacle, as indicated by the fixation ratings. Mitchell and colleagues found that when people explain future events as if they occurred in the past, they understand potential problems and implications better, because they could "see" more (1989). These findings were represented in the current experiment; when teams executed the obstacle after performing a Premortem, they understood potential problems and identified them as they began to unfold. This led teams who had conducted a Premortem to change their approach and reduce their fixation.

The Military would benefit from implanting the Premortem Method into standard planning practices, and the U.S. Army has already begun to do so (3rd BDE, USACC, 2016). The Premortem Method is better than the usual planning technique for participants operating in a time-constrained environment with emergent problems and quickly developed plans. While Veinott and colleagues examined the effects of the Premortem on confidence, no quantitative research on the effectiveness of a Premortem existed. This experiment demonstrates the value of a Premortem analysis in a Military field environment; although the manipulation was brief, outcome measures on fouls and

fixation revealed that it was more effective than previous standard operating procedures. The Premortem could easily be implemented early on in MDMP and TLP during course of action development and making a tentative plan, respectively.

The next experiment will also examine teams working in a field environment. However, individuals will be analyzing their plans without the presence or influence of their group. These plans are similar to Experiment 1 because they are team-generated. However, the plans in Experiment 2 have been carefully thought out rather than developed on the fly, and they will have a timeframe for completion of thirty days instead of thirty minutes.

Chapter 4: Experiment 2

This experiment will assess the Premortem in another field environment where pre-existing teams have already generated their own plans. However, individuals will be evaluating their plans without the presence of their groups. Instead of comparing the Premortem to a baseline, it will be compared to an evaluation step from the Scenario Planning Method; specifically, individuals will scrutinize the "Worst-Case Scenario" for their plan. This plan evaluation method was selected because it is seemingly similar to the Premortem, and people sometimes conflate the two methods; however, the Worst-Case Scenario Method does not reduce uncertainty. Worst-Case Scenario groups brainstorm reasons their plan might not be successful and identify solutions for those reasons, whereas in the Premortem groups imagine their plan failing and identify reasons and solutions. The underlying mechanisms for these plan evaluation techniques are listed in Table 4.1.

"The material contained in this chapter is in preparation for submission to a journal".

Hypotheses

H1. Participants utilizing the Premortem (failure frame and uncertainty reduction) will have a significantly larger decrease in confidence after listing reasons for plan failure.

H2. Premortem will lead to more reasons for plan failure than the Worst-Case Scenario Method. There will also be more solutions generated with the Premortem Method.

Table 4.1

Experiment 2 hypotheses

Method	**Psychological Mechanisms**	**Expected Results**
Premortem	Failure Frame + Uncertainty Reduction	More reasons and solutions generated, greater change in confidence
Worst-Case Scenario	Failure Frame	Fewer reasons and solutions generated, lesser change in confidence

Methods

Participants. Seventy-two participants were recruited for this study. They were asked to participate based on their affiliation with a group or organization that competed in MTU's month-long snow statue contest during Winter Carnival. The groups included one male-only organization, two female-only organizations, and two co-ed organizations. Participants' ages ranged 18 to 33 years old (M=20 years, SD=1.99; gender 73% male).

Five cases were excluded from analyses due to missing or incomplete data, leaving 67 cases (age *M*=20 years, *SD*=2.03; gender 75% male).

Design. This experiment is a 2x2 mixed factorial design. Participants were randomly assigned to either the Premortem Method or Worst-Case Scenario Method based on their date of birth. There were 36 participants in each condition. The independent variables examined were plan evaluation method (Premortem or Worst-Case Scenario) and knowledge of the plan. The dependent variables were number of reasons and solutions generated, and change in confidence.

Table 4.2

Questions for the Premortem and Worst-Case Scenario Methods

Premortem	Q: Imagine an outcome where your snow statue is a complete fiasco- your group has failed miserably. List as many reasons as you can as to why this happened.
Worst-Case Scenario	Q: Imagine the worst-case scenario for your group's snow statue. List as many reasons as you can as to how this scenario might happen.

Procedure. Participants received a link to participate in an online survey (Appendix B). The survey link was sent via email to a leader of each organization, and the leader forwarded the link to their constituents. The link was sent out two weeks after statue construction began; initial bulk construction was already underway. The participants were asked to rate their confidence in the plan's success on a 0-100 scale (Adams, 1957). They then evaluated their plan using either the Worst-Case Scenario Method or Premortem Method. After listing reasons their plan will/might fail, they rated

their confidence again. They then listed solutions for the problems identified. This was followed by scales measuring their subjective knowledge of the plan, commitment to the plan, and perceived likelihood of success (Appendix A). Finally, they completed a brief demographic questionnaire.

Coding Scheme. Reasons and solutions were coded separately, each divided into four corresponding categories. Two independent raters assigned each reason and solution to one of the four categories in Table 4.3; in all, 354 reasons and solutions were categorized. The raters established broad criteria for the categories; if there was disagreement or confusion between the raters, the specific reasons and categories were discussed and the category criteria were updated. Cohen's Kappa was used to determine inter-rater reliability. Ratings were conducted independently, and raters had high agreement (K=.813) (McHugh, 2012).

Table 4.3

Experiment 2 categories of reasons and solutions with descriptions.

Reasons	Solutions
External Factors (weather, timeframe, etc.) • The snow statue collapses because of all the warm weather • Statue falls apart right before judging • The snow melted • Plan is too big of a project for our group and we can't finish it in time	Change Plan/Procedure • Double shifts to make up for lost days • Redesign the smaller pieces • Everyone works double shifts as soon as it gets cold enough to work again • Carve at the last minute
Equipment • Molds aren't big enough • Can't make the barrel for the anti-air gun, due to cracking of the ice in the PVC pipe • Inefficient tools • Our supports will be too small	Equipment • Cover statue during inconvenient weather • Use a projector to sculpt to ensure we are making the right shapes • Set up tarps around statue site for easy snow removal
Leadership or Knowledge • Not enough communication between people working on it and people that planned how it is supposed to look • Statue Chair planned something incorrectly • Teams are inexperienced and do not have a leader	Leadership or Knowledge • Make sure everyone knows what the plan is for their shift • Create a GroupMe/Email to communicate to the team what is going on • Have team leader to keep group focused on task
Human Factors • Accidental breaking and/or improper carving of delicate structures • Ice detail did not go as planned • Lack of participation	Human Factors • Make participation mandatory • Balance schoolwork and statue • Provide incentives so people do better work

Results

Data were analyzed using 2x2 ANOVAs. Post-hoc t-tests were also conducted. All statistical tests used an alpha level of .05, and exact p-values are reported. P-values were corrected using the Bonferroni correction, and Greenhouse-Geisser corrections were used when Mauchly's test of sphericity indicated assumptions were violated (the F-statistic is biased and therefore invalid; there is an increased chance of a false positive, so degrees of freedom must be corrected). Results first examine specific hypotheses for each experiment, then delve into further analyses. Results did not support H1 or H2.

Table 4.4

Experiment 2 hypotheses and results

H1: Confidence, *not supported*	Individuals will have a greater change in confidence when using the Premortem Method.
H2: Reasons and Solutions, *not supported*	Individuals will generate more reasons or solutions when using the Premortem Method.

H1: Confidence. A 2x3 repeated measures ANOVA [between-subjects factors: method (Premortem, Worst-Case Scenario), within-subjects factors: time on ratings (initial confidence, confidence after listing reasons, confidence after listing solutions)] revealed a significant main effect of time on confidence, $F(2,64)=4.71$, $p=.012$; Wilk's $\Lambda=.872$, $\eta_p^2=.128$, indicating that participants' confidence ratings did change significantly across the three rating times; confidence decreased on the second rating and

increased on the third. There was no interaction effect of method and time on confidence, $F(2,64)=1.11$, $p=.336$; Wilk's $\Lambda=.966$, $\eta_p^2=.034$, indicating that any effect on confidence was not due to plan evaluation method. Separate t-tests analyzed change in confidence over time by method.

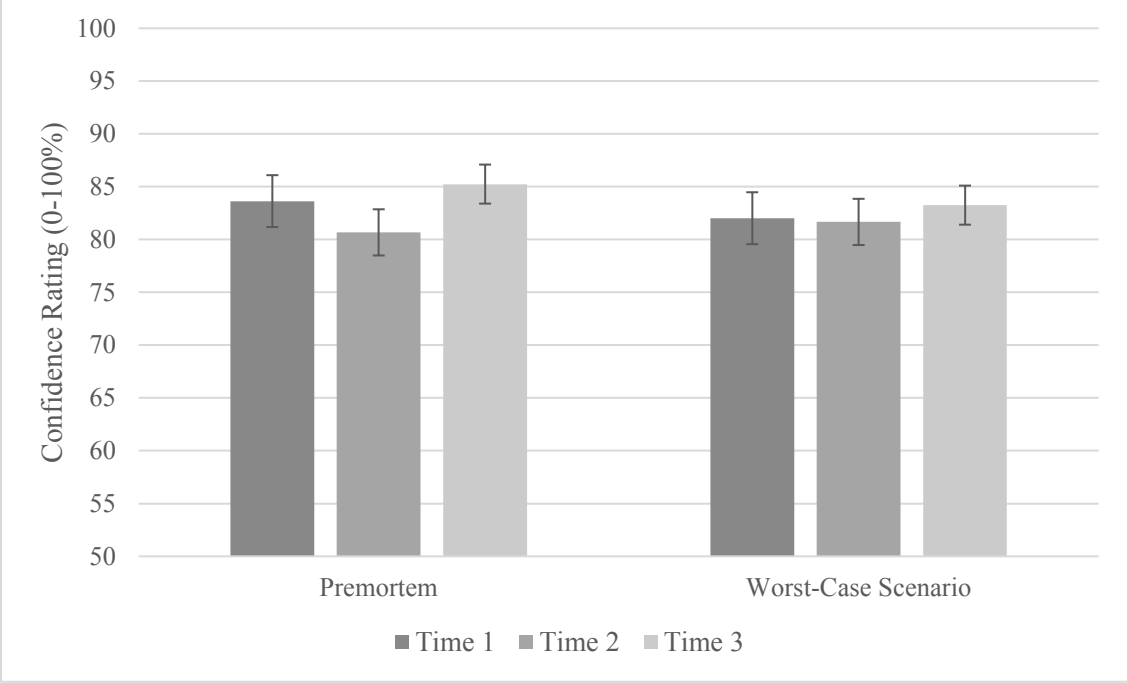

Figure 4.1. Confidence ratings by plan evaluation method and time.

For times 1 (initial confidence) and 2 (after listing reasons), there was no significant difference between change in confidence when using the Premortem Method ($M=-2.94$, $SD=9.09$) as opposed to Worst-Case Scenario Method ($M=-.33$, $SD=15.3$), $t(65)=-.848$, $p=.400$. These results indicate that confidence did not change much at all, and change in confidence was not affected by method.

For times 2 (after listing reasons) and 3 (after listing solutions), there was no significant difference between the Premortem (*M*=4.56, *SD*=10.2) and the Worst-Case Scenario (*M*=1.58, *SD*=5.41), *t*(65)=1.49, *p*=.141. These results indicate that confidence did not change much at all, and change in confidence was not affected by method.

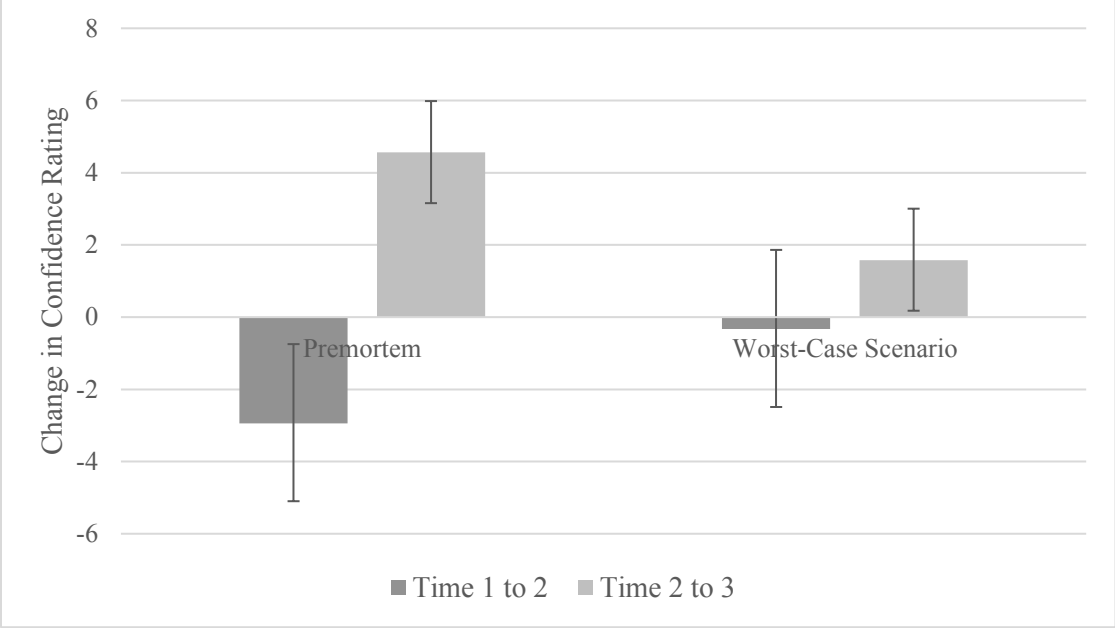

Figure 4.2. Change in confidence by plan evaluation method and time.

Knowledge. Participants were asked to rate their knowledge of their group's plan on a scale of one (no knowledge) to seven (extremely knowledgeable). Most participants were fairly knowledgeable (*M*=4.57, *SD*=1.56). Knowledge was recoded; one through four indicated low knowledge of the plan (*N*=28) and five through seven indicated high knowledge (*N*=39). Independent sample t-tests studied the effect of knowledge on confidence. Participants with higher knowledge of the plan had higher initial confidence ratings (*M*=86.4, *SD*=7.13) than those with lower knowledge (*M*=77.8, *SD*=19.3); the effect was significant, *t*(65)=2.55, *p*=.013. They also had higher confidence than lower-

knowledge participants in their second and third confidence ratings, but the effects were not significant.

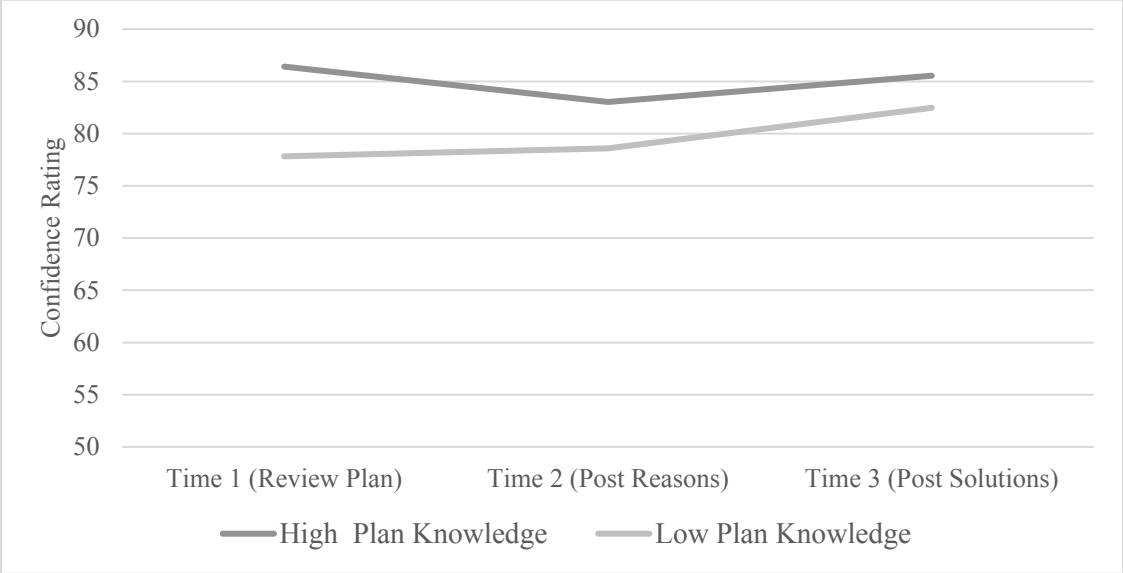

Figure 4.3. Change in confidence by knowledge of plan and time.

H2: Reasons and Solutions by Method. A 2x2 MANOVA [between-subject factors: method (Premortem, Worst-Case Scenario), within-subject factors: reasons and solutions] was conducted to examine number of reasons and number of solutions by method. Results found no main effect of method on number of reasons and solutions, $F(2,64)=1.02$, $p=.368$, Wilk's $\Lambda= .969$, $\eta_p^2=.031$. Participants using the Premortem Method did not generate statistically more reasons, $F(1,66)=2.01$, $p=.161$, $\eta_p^2=.030$, or solutions, $F(1,66)=.646$, $p=.425$, $\eta_p^2=.010$ than participants using the Worst-Case Scenario Method.

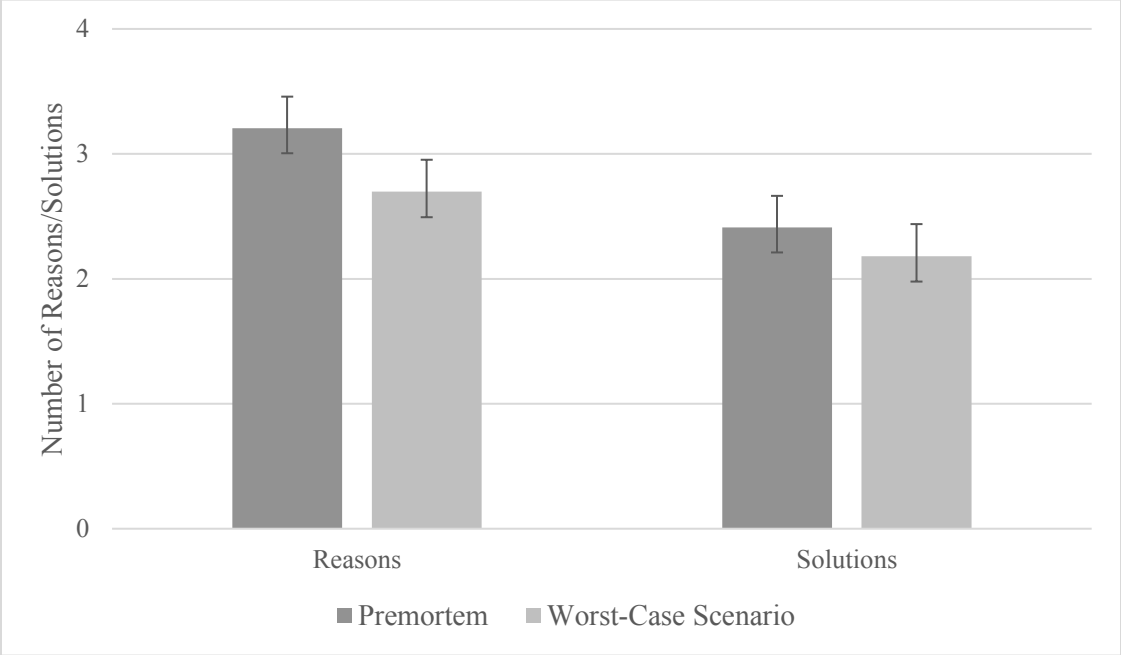

Figure 4.4. Number of reasons and solutions by plan evaluation method.

Knowledge. A 2x2 MANCOVA [between-subject factors: method (Premortem, Worst-Case Scenario), within-subject factors: reasons and solutions; covariate: knowledge group] showed no main effect of method, $F(2,63)=1.25$, $p=.293$; Wilk's $\Lambda= .962$, $\eta_p^2=.038$, indicating that regardless of their level of knowledge, participants did not generate more reasons or solutions when using the Premortem Method. However, there was an statistically significant effect of knowledge on number of reasons and solutions generated, $F(2,63)=3.42$, $p=.039$, Wilk's $\Lambda= .902$, $\eta_p^2=.098$. Tests of between-subjects effects found an effect of knowledge on number of reasons, $F(1,66)=6.15$, $p=.016$, $\eta_p^2=.088$. There was also an effect of knowledge on number of solutions, $F(1,66)=5.31$, $p=.024$, $\eta_p^2=.077$. Participants with a higher subjective knowledge of the plan generated more reasons and solutions.

Independent sample t-tests studied the effect of knowledge on number of reasons and solutions. Participants with higher knowledge of the plan generated more reasons (M=3.31, SD=1.59) than those with lower knowledge (M=2.46, SD=1.17); the effect was statistically significant, $t(65)=2.38$, $p=.020$. Participants with higher knowledge also generated more solutions (M=2.56, SD=1.23) than those with lower knowledge (M=1.93, SD=.979); the effect was statistically significant, $t(65)=2.26$, $p=.027$.

Types of Reasons. Participants using the Premortem Method generated 110 reasons the plan failed; they were categorized and raters had high agreement (K=.885). Participants using the Worst-Case Scenario Method generated 89 reasons that a worst-case scenario might occur; they were categorized and raters had substantial agreement (K=.609). A chi-squared test was performed to examine the relationship between reason categories across methods; the effect was significant, $X^2(3)=8.88$, $p=.031$. This indicates that the distribution of reasons was different across the two methods; participants using the Premortem Method focused more on leadership, knowledge, and human factors (internal factors), whereas participants using the Worst-Case Scenario Method focused more on external factors.

Table 4.5

Experiment 2 crosstabulation for types of reasons

		Condition		Total
		Premortem	Worst-Case Scenario	
Type of Reason	External Factors	36 (33%)	45 (51%)	81 (41%)
	Equipment	11 (10%)	11 (12%)	22 (11%)
	Leadership, Knowledge	25 (23%)	16 (18%)	41 (21%)
	Human Factors Errors	38 (35%)	17 (19%)	55 (27%)
	Total	110	89	199

Types of Solutions. Participants using the Premortem Method generated 89 solutions or ways to strengthen the plan; they were categorized and raters had high agreement (K=.814). Participants using the Worst-Case Scenario Method generated 82 solutions or ways to strengthen the plan; they were categorized and raters had high agreement (K=.895). A chi-squared test was performed to examine the relationship between solution categories across plan evaluation methods; the effect was statistically significant, $X^2(3)=13.40$, $p=.004$. This indicates that participants were coming up with different types of solutions with the two different methods; participants using the Premortem Method focused more on leadership, knowledge, and human factors whereas participants using the Worst-Case Scenario Method focused more on changing the plan or procedure.

Table 4.6

Experiment 2 crosstabulation for types of solutions

		Condition		Total
		Premortem	Worst-Case Scenario	
Type of Solution	Change Plan/Procedure	21 (26%)	35 (48%)	56 (36%)
	Equipment	15 (18%)	15 (20%)	30 (19%)
	Leadership, Knowledge	31 (38%)	20 (27%)	51 (33%)
	Human Factors Solutions	15 (18%)	3 (4%)	18 (11%)
	Total	82	73	155

Discussion

Hypotheses for Experiment 2 were not supported in general. There was no effect of method on confidence, number of reasons, or number of solutions. However, the pattern of reasons and solutions differed across methods. There are several variables that could have contributed to these findings.

Knowledge of the plan is crucial for plan evaluation. If a team member does not have sufficient knowledge or understand the plan, they cannot be expected to adequately assess it. Before data collection, snow statues were halfway complete and leaders were asked to ensure everyone was familiar with the plan; however, analyses of the types of reasons and solutions generated suggest that many group members were unaware of the full plan. Knowledge has an effect on initial confidence; if group members have less knowledge of the plan, they have less initial confidence. Furthermore, participants with

less knowledge generated fewer reasons and solutions; they were unable to think about what might go wrong with the plan because they did not fully understand what the plan entailed.

Snow statue plans are likely to go sideways in unfavorable weather conditions. During data collection, weather at MTU was unseasonably warm. If the temperature rises above 45 degrees Fahrenheit or dips below -25 degrees Fahrenheit, organizations are not permitted to work on their snow statues due to safety risks. Several days during the data collection period were too warm for groups to make physical progress on their statues. This is reflected in the data; 41% of reasons participants generated across both methods were about external factors such as weather. If unfavorable weather had not been a factor during this timeframe, participants might have listed different, more insightful reasons the plan will/might fail.

The Premortem has typically been conducted face-to-face with a facilitator. This experiment collected data through an anonymous survey; participants were told to spend two minutes each generating reasons and solutions. However, the lack of control and anonymity might have contributed to participants spending less time thinking through their plan and potential shortcomings.

While the present hypotheses were not supported, a more important finding surfaced after examining the types of reasons and solutions. The significant difference in distribution demonstrates that the underlying mechanisms of the Premortem Method and Worst-Case Scenario Method impacted the types of reasons and solutions participants

generated. Only one sentence differed between the two conditions; yet, that manipulation caused participants to frame the problem differently. Framing changes how people search for information; it changes how the problem is perceived by altering which aspects of the problem are prominent (Koehler, 1991). Participants using the Worst-Case Scenario Method fixated on one scenario, one frame; many participants fixated on their statue falling over due to poor weather conditions. However, when participants used the Premortem Method, they broadened their conditional reference frame and viewed failure resulting from several possible reasons, instead of just one. Therefore, they considered downfalls in human factors, leadership, and knowledge as reasons the plan would fail. Attempting to explain future events as if they happened in the past helps people understand potential problems and implications better (Mitchell, Russo, & Pennington, 1989). Participants using the Worst-Case Scenario Method predominately suggested solutions related to changing the plan or procedures used, whereas participants using the Premortem Method generated solutions related to leadership and human factors.

Attribution theory is one explanation for the fundamental differences between the types of reasons and solutions generated. Heider (1958) proposed that people have a strong motivation to form a coherent understanding of the world and a need to control the environment around them. When participants undergo the Premortem manipulation, they are informed that the plan completely failed. Per attribution theory, this creates a desire to understand why the plan failed; the underlying motivations change. When participants imagine their worst-case scenario, they do not have the same craving to find answers.

They are imagining one potential outcome, instead of imagining an end state and generating all the possible outcomes that could lead to that end state.

Experiment 2 examined individuals analyzing their own group plans without group discussion. The evaluation was conducted online over a short period of time, and did not support the present hypotheses; however, the types of reasons and solutions generated suggest that the underlying psychological mechanisms of the Premortem Method and Worst-Case Scenario Method are fundamentally different, and require further analyses. The next experiment once more examines the Premortem and Worst-Case Scenario Methods; this time, the exercises will be facilitated and groups will use a plan with which they are unfamiliar.

Chapter 5: Experiment 3

One explanation for the lack of difference in Experiment 2 between the two planning methods was that they were done individually, outside of a group. Experiment 3 sought to directly compare the Premortem Method with the Worst-Case Scenario Method, as was done in Experiment 2; however, in this experiment, participants will be working in groups of 3-5, they will be in a conference room setting, and they will not be using their own plans. There will also be an individual condition where participants will conduct a Premortem with a facilitator, but without group discussion.

Hypotheses

H1. Participants using the Premortem Method will have a greater change in confidence than those using the Worst-Case Scenario Method. Participants using the Premortem Method with a group discussion will have the greatest change in confidence.

H2. Participants in groups will generate more reasons and solutions than individuals, and participants using the Premortem Method will generate more reasons and solutions than those using the Worst-Case Scenario Method.

Methods

Participants. Eighty-seven participants were recruited. They were all enrolled in an Introductory to Psychology course and received credits towards their class grade for their participation in this study. There were 31 participants in the Premortem group condition (seven groups), 27 in the Scenario group condition (six groups), and 29 in the

---

"The material contained in this chapter is in preparation for submission to a journal".

Premortem individual condition. Five participants (one Premortem group) were dropped from this study due to four out of the five group members lacking a proficiency in writing in English. Of the remaining 82 participants, the average ages ranged 18 to 26 years (M=19.5, SD=1.51; gender 65.9% male).

Design. This experiment is a 3x2 mixed factorial design. The independent variable is method (Premortem Method, Worst-Case Scenario Method, or Premortem Individual Method). The Premortem and Worst-Case Scenario Methods were compared, and the Premortem and Premortem Individual Methods were compared. The dependent variables are confidence and number of reasons and solutions.

Procedure. Students were recruited from an Introductory to Psychology course, and signed up in groups of up to five participants. Timeslots were randomly assigned to the Premortem or scenario method until each condition had at least six groups; the remaining participants were assigned to the individual Premortem Method. The entire experiment was conducted over 15 days.

Premortem Group. The group conditions all contained no fewer than three and no more than five people. Participants were seated in a room with the facilitator. After signing informed consent forms, participant received a packet and were told to follow instructions at the bottom of the pages. Before beginning, each group was given a brief explanation of Norovirus:

Norovirus is a highly contagious RNA-type virus, like seasonal influenza, that gives people flu-like symptoms; it is difficult to create a vaccination for this virus.

Participants familiarized themselves with the Norovirus lockdown plan (Appendix B) and rated their confidence in the plan. Once everyone had completed these initial steps, the facilitator introduced the Premortem manipulation:

> *Imagine 6 months into the future: there was a need for this plan to be implemented. Michigan Tech implemented the plan, and it was a complete fiasco. It was a catastrophic failure. Take the next two minutes to write down as many reasons as you can think of as to why this plan failed.*

The participants went through the steps of the Premortem process: two minutes to write down reasons why the plan failed; a group discussion of the reasons generated; second confidence and understanding ratings; two minutes to generate solutions or ways to strengthen the plan; a group discussion of solutions generated; and final confidence and understanding ratings. They were also asked to rate their knowledge of MTU policy, Norovirus, and lockdown plans. Finally, they completed a demographic questionnaire. Upon completion, participants were de-briefed.

Premortem Individual. The individual condition was the same as the Premortem group condition; however, there was no discussion of reasons and solutions generated.

Worst-Case Scenario Group. The Worst-Case Scenario groups were the same as the Premortem group condition in the beginning; they read through the plan and completed initial confidence and understanding ratings. However, instead of being told to imagine the plan failing, they were instructed to imagine the worst-case scenario for this plan:

> *Now, imagine the worst-case scenario that might happen if this plan were to be implemented. Write it down, and take 2 minutes to write down as many reasons as you can as to how that worst-case scenario might happen.*

The rest of the evaluation was the same as the Premortem Method; a group discussion of the reasons generated; second confidence and understanding ratings; two minutes to generate solutions or ways to strengthen the plan; a group discussion of solutions generated; and final confidence and understanding ratings. They were also asked to rate their knowledge of MTU policy, Norovirus, and lockdown plans. Finally, they completed a demographic questionnaire. Upon completion, participants were de-briefed. The only difference between the Premortem manipulation and the Worst-Case Scenario manipulation was the phrasing and written instructions for generating reasons. Premortem eliminates uncertainty and describes the problem as "plan failure." The Worst-Case Scenario method does not eliminate uncertainty and fixates users on one worst-case scenario that may or may not imply plan failure.

Table 5.1

Premortem, Worst-Case Scenario, and Premortem Individual comparison.

Plan Evaluation Method	Number of Participants	Uncertainty Reduction?	Failure Frame?	Group?
Premortem	26	x	x	x
Worst-Case Scenario	27		x	x
Premortem Individual	29	x	x	

Coding Scheme. Reasons and solutions were each divided into four corresponding categories. The reasons for failure are Quarantine, Education, Security, Resources, and Other. Solutions to the reasons are labeled the same. Two independent raters assigned each reason and solution to a category; 538 reasons and solutions were categorized. The raters established broad criteria for the categories; if there was disagreement or confusion between the raters, the specific reasons and categories were discussed and the category criteria were updated. Cohen's Kappa was used to determine inter-rater reliability; the raters had high agreement (K=.817) (McHugh, 2012). The categories are represented in Table 5.2 with examples from the data; further analyses of categories are discussed in the Results section.

Table 5.2

Experiment 3 categories of reasons and solutions with descriptions.

Reasons	Solutions
Spread and Contamination • Healthy people being stuck on campus with sick people • People start to panic and leave campus while infected	**Quarantine** • Once quarantined, don't let people leave unless proved clear • Provide information to students on how to prevent spread.
Education • Students fall way behind on school year after outbreak • An immediate switch to online schooling might not go well • Professors get ill	**Education** • Give instructors online resources • Classes are not priority; health is priority • All classes shut down until further notice
Security • Proper access to campus could be forged and used to cause terrorism • The security guards themselves got it and passed it on	**Security** • Specify how/where entrance to campus will be blocked • Require two photo IDs and MTU ID for any access
Resources • Failures in food safety • Water supply is infected	**Resources** • Pass out masks, bottled water, gloves • Bring in outside doctors/nurses
Other • The plan has no end/recovery phase • People didn't get enough information and didn't take it seriously	**Other** • Tell everyone the plan ahead of time • Offer incentives to ensure students follow safety protocol

Results

Data were analyzed using 3x3 ANOVAs to examine the effects of plan evaluation method on confidence, reasons, and solutions. Results partially supported H1 and H2.

Table 5.3

Experiment 3 hypotheses and results.

H1: Confidence, *partially supported*	Participants using the Premortem Method will have a greater change in confidence than those using the Worst-Case Scenario Method.
H2: Reasons and Solutions, *partially supported*	Participants in groups will generate more reasons and solutions than individuals, and participants using the Premortem Method will generate more reasons and solutions than those using the Worst-Case Scenario Method.

H1: Confidence. A 3x3 repeated measures ANOVA [between-subjects factors: method (Premortem, Scenario, Premortem individual), within-subjects factors: confidence (initial confidence, confidence after listing reasons, final confidence)] revealed a significant main effect of confidence, $F(2,78)=112$, $p<.001$, Wilk's $\Lambda=.259$, $\eta_p^2=.741$, indicating that participants' confidence ratings changed significantly across the three rating periods. Furthermore, there was a marginally significant interaction of confidence and method, $F(4,156)=2.17$, $p=.075$, Wilk's $\Lambda=.898$, $\eta_p^2=.053$. Mauchly's test indicated the assumption of sphericity had been violated [$X^2(2)=6.68$, $p=.036$]. Degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity

(ε=0.92). A main effect of confidence, $F(1.85,146)=115$, $p<.001$, $\eta_p^2=.594$, was qualified by an interaction between confidence and method, $F(3.70,146)=2.83$, $p=.031$, $\eta_p^2=.067$, indicating that participants using the Premortem Method with a group had a more substantial change in confidence than the other two methods across the three rating periods (Figures 5.1a and 5.1b). Separate t-tests analyzed the changes in confidence across Rating Times 1 and 2 and Times 2 and 3.

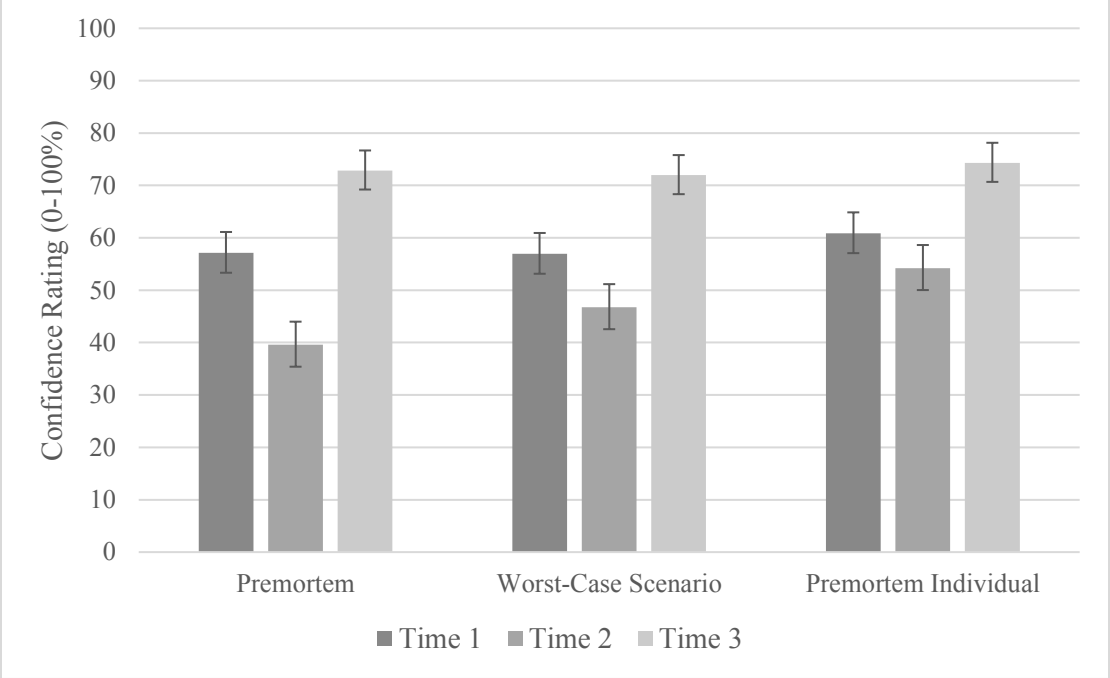

Figure 5.1a. Confidence ratings by plan evaluation method and time.

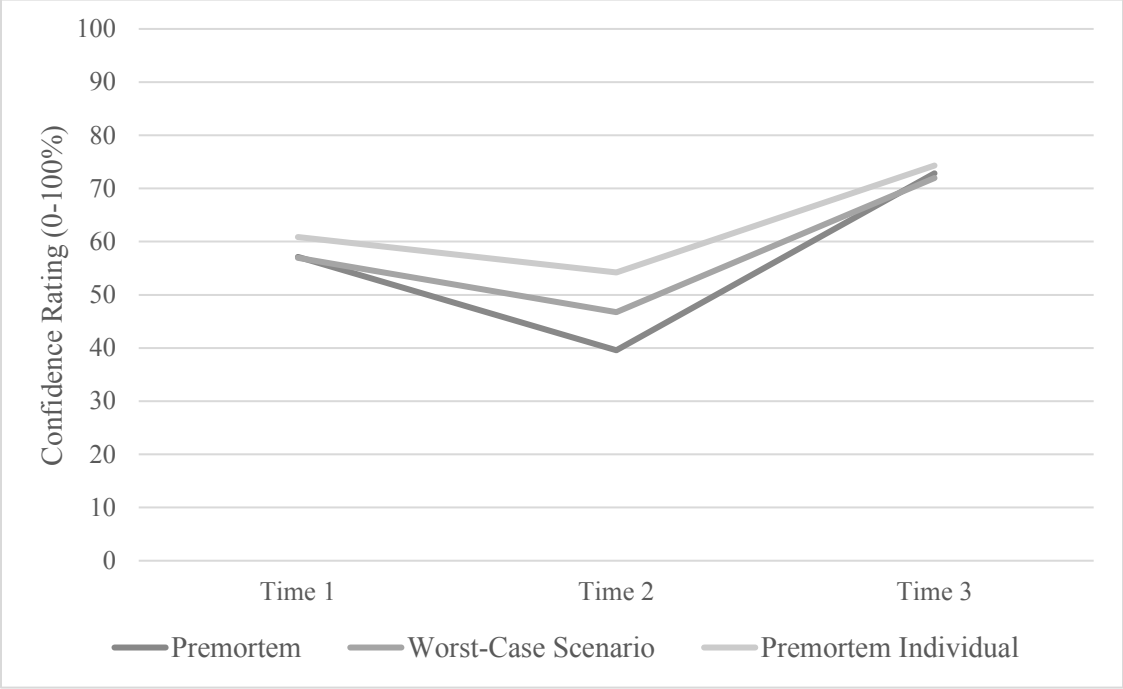

Figure 5.1b. Confidence ratings by plan evaluation method and time.

For Times 1 (after getting familiar with the plan) and 2 (after listing reasons), there was no significant difference between Premortem (M=-17.5, SD=16.4) and Worst-Case Scenario (M=-10.19, SD=14.7), $t(51)$=-1.717, p=.092. There was, however, a statistically significant difference between Premortem and Premortem Individual, $t(53)$=-2.28, p=.027. These results indicate that the Premortem Method is superior to the Premortem Individual Method in reducing overconfidence. An important factor here is the group dynamic; with the Premortem Method and Worst-Case Scenario Method, Rating Time 2 occurred after reasons had been discussed with the group. With the Premortem Individual Method, Rating Time 2 occurred after individuals wrote down their own reasons for failure, with no group discussion.

For Times 2 (after listing reasons) and 3 (after listing solutions), there was no significant difference between the Premortem (M=33.3, SD=16.7) and the Worst-Case Scenario (M=25.2, SD=17.6), $t(51)=1.71$, $p=.094$. There was, however, a statistically significant difference between Premortem and Premortem Individual, $t(53)=3.01$, $p=.004$. These results indicate that the Premortem Method is superior to the Premortem Individual Method in raising confidence after lowering it.

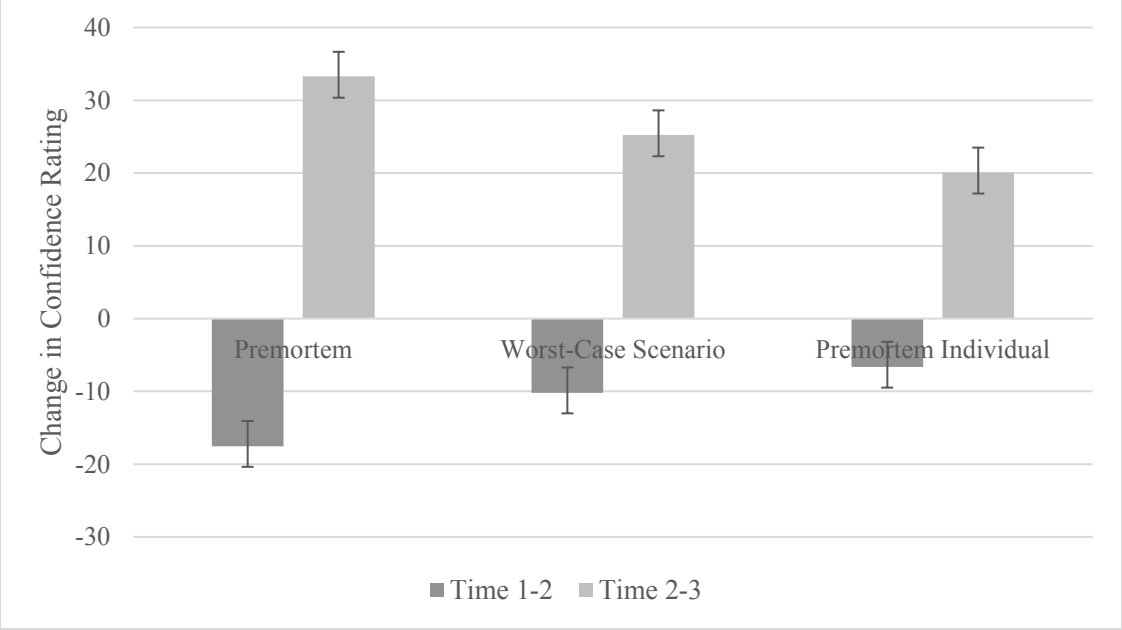

Figure 5.2. Change in confidence by plan evaluation method and time.

Types of Reasons. Participants using the Premortem Method generated 98 reasons for plan failure; they were categorized, and raters had substantial agreement (K=.787). Participants using the Worst-Case Scenario Method generated 52 reasons that a worst-case scenario might occur; they were categorized, and raters had high agreement (K=.815). Participants using the Premortem Individual Method generated 98 reasons for plan failure; they were categorized, and raters had high agreement (K=.890). A chi-

squared test was performed to examine the relationship between reason categories across methods; the effect was significant, $X^2(8)=19.3$, $p=.013$. This indicates that the distribution of reasons was different across the three methods; participants using the Worst-Case Scenario Method focused mainly on the quarantine. Participants in the Premortem group condition generated more reasons pertaining to resources than the other two conditions did; furthermore, participants in the Premortem individual condition focused more on security than the other two conditions did.

Table 5.4

Experiment 3 crosstabulation for types of reasons.

		Condition			Total
		Premortem (N=26)	Worst-Case Scenario (N=27)	Premortem Individual (N=29)/	
Type of Reason	Quarantine	57 (58%)	39 (75%)	60 (61%)	156 (63%)
	Education	5 (5%)	2 (3%)	3 (3%)	10 (4%)
	Security	8 (8%)	6 (12%)	18 (18%)	32 (13%)
	Resources	16 (16%)	0 (0%)	5 (5%)	21 (8%)
	Other	12 (12%)	5 (10%)	12 (12%)	29 (12%)
	Total	98	52	98	248

Types of Solutions. Participants using the Premortem Method generated 114 solutions or ways to strengthen the plan; they were categorized, and raters had high agreement (K=.834). Participants using the Worst-Case Scenario Method generated 90 solutions or ways to strengthen the plan; they were categorized, and raters had substantial

agreement (K=.723). Participants using the Premortem Individual Method generated 86 solutions or ways to strengthen the plan; they were categorized, and raters had high agreement (K=.831). A chi-squared test was performed to examine the relationship between solution categories across methods; the effect was not significant, $X^2(8)=6.36$, $p=.608$. This indicates that the percentage of solutions within each category did not vary significantly across the three methods.

Table 5.5

Experiment 3 crosstabulation for types of solutions.

		Condition			Total
		Premortem (N=26)	Worst-Case Scenario (N=27)	Premortem Individual (N=29)	
Type of Solution	Quarantine	50 (44%)	34 (38%)	38 (44%)	122 (42%)
	Education	11 (10%)	10 (11%)	8 (9%)	29 (10%)
	Security	13 (11%)	6 (7%)	13 (15%)	32 (11%)
	Resources	27 (24%)	27 (30%)	16 (18%)	70 (24%)
	Other	13 (11%)	13 (14%)	11 (13%)	37 (12%)
	Total	114	90	98	290

Reasons and Solutions by Method. A multivariate ANOVA [between-subjects factors: method (Premortem, Worst-Case Scenario, Premortem Individual), within-subjects factors: reasons and solutions] was conducted to examine the effects of method on number of reasons and solutions generated. Results found an effect of method, $F(2,78)=10.3$, $p<.001$, Wilk's $\Lambda=.627$, $\eta_p^2=.208$, indicating that participants using the

Premortem Method with a group generated more reasons and solutions than participants using the Worst-Case Scenario and Premortem individual methods. The interaction was statistically significant, $F(2,79)=9.36$, $p<.001$. Between-subjects analysis showed a significant effect of method on number of reasons generated, $F(2,81)=14.6$, $p<.001$, $\eta_p^2=.270$, indicating that participants in the Premortem conditions generated more reasons than participants using the Worst-Case Scenario method. Furthermore, there was also a significant of method on number of solutions, $F(2,81)=7.56$, $p=.001$, $\eta_p^2=.161$, indicating that participants using the Premortem Method in a group generated the most solutions, followed by the Worst-Case Scenario method. Participants using the Premortem Individual method generated the least amount of solutions (Figure 5.3).

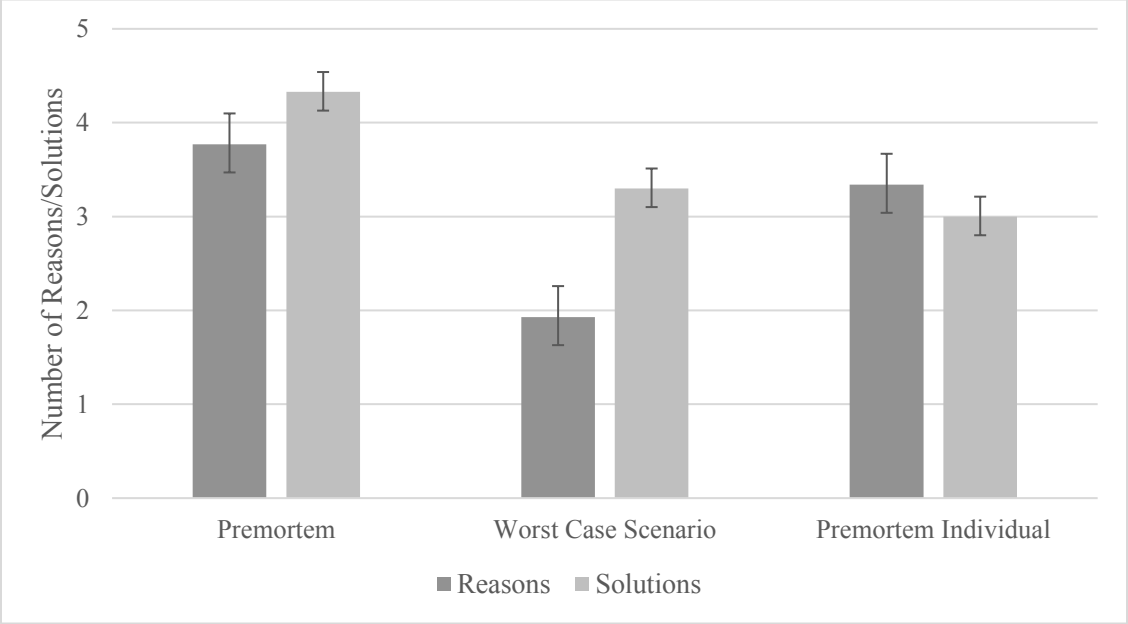

Figure 5.3. Number of reasons and solutions by plan evaluation method.

Post-hoc analyses indicated that fewer reasons were generated with the Worst-Case Scenario Method ($M=1.93$, $SD=1.04$) than the Premortem Method ($M=3.77$,

SD=1.53); the effect was statistically significant, $t(51)=5.15$, $p<.001$. Furthermore, the Worst-Case Scenario Method was less effective than the Premortem Individual Method (*M*=3.34, *SD*=1.11); the effect was statistically significant, $t(54)=4.46$, $p<.001$. This indicates that participants were generating significantly fewer reasons when using the Worst-Case Scenario method rather than the Premortem Method or Premortem Individual Method. There was no significant difference in number of reasons generated between the Premortem and Premortem Individual, $t(53)=1.11$, $p=.274$.

Table 5.6

Group statistics for number of reasons

	Method	N	Mean	Std. Deviation	Std. Error Mean
Number of Reasons	Premortem	26	3.77	1.53	.300
	Worst-Case Scenario	27	1.93	1.04	.199
	Premortem Individual	29	3.34	1.32	.245

More solutions were generated using the Premortem Method (*M*=4.38, *SD*=1.68) than the Worst-Case Scenario Method (*M*=3.33, *SD*=1.11); the difference was statistically significant, $t(51)=2.70$, $p=.009$. The Premortem Method was also more effective than the Premortem Individual Method (*M*=3.07, *SD*=1.10); the difference was statistically significant, $t(53)=3.48$, $p=.001$. Participants did not generate significantly more solutions when using the Worst-Case Scenario Method than the Premortem Individual Method, $t(54)=.895$, $p=.375$. Overall, this indicates that participants identified

significantly more solutions or ways to strengthen the plan when using the Premortem Method than the Worst-Case Scenario Method. Furthermore, while participants generated a similar amount of reasons using the Premortem and Premortem Individual Methods, participants generated the fewest amount of solutions when using the Premortem Individual Method.

Table 5.7

Group statistics for number of solutions

	Method	N	Mean	Std. Deviation	Std. Error Mean
Number of Solutions	Premortem	26	4.38	1.68	.329
	Worst-Case Scenario	27	3.33	1.11	.214
	Premortem Individual	29	3.07	1.10	.204

Understanding. A 3x3 repeated measures ANOVA with a within-subjects factor of understanding of the plan (initial understanding, understanding after listing reasons, final understanding) and a between-subjects factors of method (Premortem, Worst-Case Scenario, Premortem Individual) revealed a significant main effect of understanding, $F(2,78)=6.17$, $p=.003$, Wilk's $\Lambda=.863$, $\eta_p^2=.137$, and an interaction of understanding and method, $F(4,156)=4.41$, $p=.002$, Wilk's $\Lambda=.807$, $\eta_p^2=.102$. Mauchly's test indicated the assumption of sphericity had been violated [$X^2(2)=29.1$, $p<.001$]. Degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\varepsilon=0.762$). There was a main effect of understanding, $F(1.53,120)=8.90$, $p=.001$, $\eta p2=.101$, indicating that

participants' level of understanding of the situation changed throughout the exercise. Furthermore, there was a statistically significant interaction between understanding and method, $F(3.05,120)=6.04$, $p=.001$, $\eta p2=.133$, indicating that method affected understanding; specifically, participants' final understanding measure decreased in the Premortem Individual condition whereas it increased in the other two conditions.

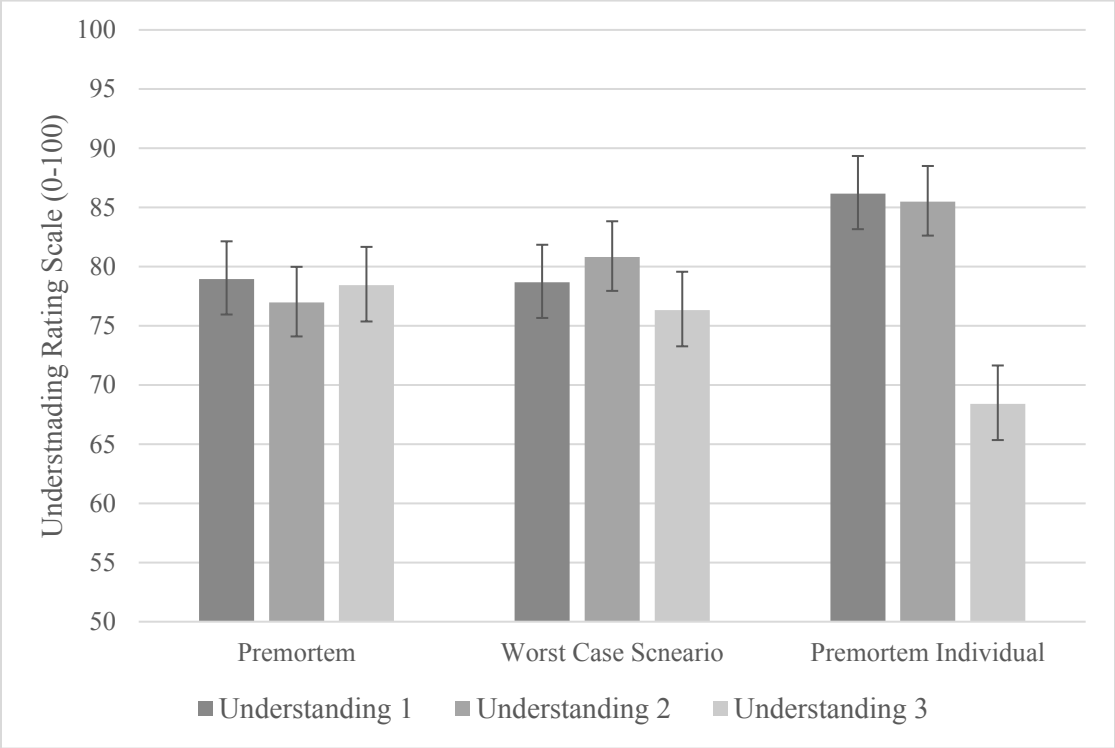

Figure 5.4. Understanding ratings by plan evaluation method and time.

Discussion

Overall, the present results partially supported the hypotheses.

H1. *Participants using the Premortem Method will have a greater change in confidence than those using the Worst-Case Scenario Method.*

Hypothesis 1 was partially supported by the present findings. The Premortem Method was more effective at reducing overconfidence than the Worst-Case Scenario Method; however, the Premortem Individual Method produced the least change in confidence and the highest confidence scores. The failure frame and prospective hindsight mechanism alone do not reduce overconfidence; group discussion is an important factor. Nonetheless, the findings on change in confidence when using the Premortem method replicate previous research; the Premortem is the most effective evaluated method for producing a change in confidence across three rating periods (Veinott, Klein, & Wiggins, 2010).

H2. *Participants in groups will generate more reasons and solutions than individuals, and participants using the Premortem Method will generate more reasons and solutions than those using the Worst-Case Scenario Method.*

Hypothesis 2 was also partially supported. Participants in the Premortem group generated significantly more reasons and solutions than participants using the Worst-Case Scenario Method and the Premortem Individual Method. However, the Worst-Case Scenario Method produced significantly more solutions than the Premortem Individual Method. Participants generated significantly more reasons when using the Premortem and Premortem Individual Methods as opposed to the Worst-Case Scenario Method. Again, these results can be partially attributed to the group dynamic: participants in groups came up with more solutions, because the group discussion brought up many more reasons. Participants conducting a Premortem alone, without group discussion, were only generating solutions for their own reasons.

The types of reasons and solutions generated varied across methods. This indicates that the underlying mechanisms of the methods affected how participants approached problems and created solutions. The group and individual Premortem conditions yielded similar types of reasons, because the manipulation was the same. The group and individual conditions had similar types of solutions, but participants using the Premortem Individual Method generated far fewer solutions than participants using both the Premortem Method and Worst-Case Scenario Method. Participants using the Worst-Case Scenario Method generated far fewer reasons the worst-case scenario might occur. They were in a failure frame; however, they were focused on only one scenario. Premortem allowed participants to explore myriad worst-case scenarios instead of fixating on one and trying to think of all the ways one scenario might occur. While the Worst-Case Scenario Method might be useful in identifying and preventing one high-risk undesirable scenario, the Premortem Method can identify several poor outcomes and more reasons a plan might go sideways.

Previous research suggests that "groupthink" mentality can prevent people from seeing flaws in a plan (Puncochar & Fox, 2004). People tend to believe that the group performance is better than their individual performance; when one group member speaks, it could inhibit others from sharing their own thoughts (Plous, 1995). In the present Experiment, when participants used the Premortem Method and Worst-Case Scenario Method, they were asked to discuss their reasons and solutions with the group. Each group member spoke about what they wrote down, and the facilitator transcribed all reasons and solutions so everyone could see them. In several cases, participants wrote

down more reasons or solutions than they shared with the group. As evidenced in the results, group discussion helped participants generate more solutions; often they came up with solutions for other participants' reasons in addition to their own. These findings support literature on brainstorming; nominal groups that involve individual brainstorming followed by group discussion are the most effective (Bouchard, 1969). When participants analyzed the same plan using the Premortem Individual Method, they generated far fewer solutions. These results imply that while a group discussion might suppress people from sharing all their reasons for failure, the benefit outweighs the cost, because they still generated more ways to improve the plan than individuals working alone. The group component to plan evaluation is important for strengthening a plan.

When people explain sure outcomes, they generate more reasons those outcomes will occur (Mitchell, Russo, & Pennington, 1989). Participants using the Premortem Method and Premortem Individual Method generated far more reasons the plan would fail, because it was a sure outcome. Participants generating reasons the worst-case scenario might occur did not come up with as many reasons, because the end state was hypothetical. The types of reasons were also distributed differently across the five categories; as was seen in Experiment 2, there was a fundamental difference in how participants approached the problem. As Mitchell and colleagues suggested, participants were using different strategies to explain sure outcomes (1989). Experiment 3 expanded on Experiment 2 while comparing the effect of a group dynamic in the Premortem process. Results indicate that while a group is important for generating solutions for

problems and ways to strengthen a plan, the underlying mechanisms of the Premortem Method are maintained regardless of group presence or type of plan.

Chapter 6: Overall Discussion

The present research extends research on plan evaluation in several ways. First, it extends the study of the Premortem plan evaluation technique to a field setting and a short time frame (5-10 minutes) process, and demonstrates that it is just as effective. This is the first time this technique was qualitatively and quantitatively validated in a field setting and under a range of time constraints. Previously, a Premortem took about an hour to complete; Experiment 1 showed it to be effective in a few minutes. Furthermore, the Premortem was typically conducted in teams; Experiments 2 and 3 attempted to validate the method with individuals.

Experiment 1 sought to validate the Premortem in a shorter timeframe with team-generated and executed plans. Previously, there were no outcome measures for the Premortem. Military planning and decision making relies on gathering as much relevant information as possible; there is rarely certainty in Military command decisions. Collaboration is important; understanding the current state of the environment and imagining a desired end state helps leaders determine the best course of action (Perez, 2011). With so many factors to consider, operating in a field environment with many unknowns presents a challenge and opportunity for growth and development in planning and decision making skills.

Many plans are made in collaboration with other leaders, which can present a groupthink mentality and prevent key leaders from seeing weaknesses or shortcomings in their plans (Puncochar & Fox, 2004). Groupthink Mitigation (GTM) can alleviate these

problems by fostering divergent thinking and including all team members' perspectives before selecting a course of action. If group members write down their thoughts before group discussion, they will be less likely to succumb to groupthink, and will be able to examine the plan more effectively (3rd BDE, USACC, 2016). A Premortem analysis of the plan can reduce groupthink; most importantly, the group discussion of why a plan failed helps leaders maintain their individual perspective while sharing in the viewpoints of others.

Recent publications suggest implementing the Premortem into the MDMP process; specifically, leaders are encouraged to conduct a Premortem prior to selecting a course of action (Tradoc G2 Operational Environment Enterprise, 2016). The Premortem process has been recognized by the Military as an effective tool for plan evaluation, but is not yet ubiquitous. Aside from research conducted by Veinott and colleagues, the present research is the only quantitative evidentiary support for the validity of the Premortem plan evaluation method. Furthermore, Experiment 1 validates this process in a Military field environment. The effectiveness of the Premortem on reduction in fouls and fixation provides a compelling argument that the process should be ingrained in the MDMP process and utilized by smaller units in time-constrained field settings. Specifically, the MDMP and TLPs could be enhanced by implementing a Premortem early in the planning process. In MDMP, a Premortem could be conducted in the beginning of step 3 (course of action development). If a smaller element is using TLPs, a Premortem could be conducted during step 1 (receive and analyze the mission). Conducting a Premortem as

early as possible can drive the planning process by helping leaders elaborate and improve options from the beginning, instead of satisficing down the road (Klein, 2008).

Further experimentation revealed the importance of the underlying psychological mechanisms in the Premortem technique. Experiment 2 directly compared the Premortem Method and the Worst-Case Scenario Method. While the number of reasons and solutions generated did not significantly differ, the distribution of reasons and solutions across the four different categories varied significantly across the two plan evaluation methods. The Premortem Method, utilizing the prospective hindsight mechanism, generated far more reasons and solutions pertaining to leadership, knowledge, and human factors. Attribution theory can possibly explain the difference in types of reasons and solutions generated; when uncertainty is reduced, people will try to generate causal reasons why something happened so it makes sense to them (Heider, 1958).

Judging the value of reasons and solutions will help us determine if prospective hindsight is helping people generate superior plans (Mitchell, Russo, & Pennington, 1989). The current research did not qualitatively judge the value of each reason and solution generated. However, reasons pertaining to knowledge, leadership, and human factors are malleable and can be controlled. They are therefore subjectively viewed as more important; these are issues that are often overlooked, but can be manipulated to strengthen the plan. Reasons related to external factors and equipment have less to do with the plan and group performance; changes can be made to the plan or procedure, but ultimately, underlying problems with the plan are not addressed. Consequently, the

reasons and solutions generated using the Premortem Method are superior to reasons generated using the Worst-Case Scenario Method.

The number of reasons and solutions generated in Experiment 2 was hypothesized to be greater when using the Premortem Method. This hypothesis was not supported and was re-examined in Experiment 3. The effects of group dynamics were also examined. Previous research suggests that people think their group performance is superior to their individual performance (Plous, 1995) and that groupthink can prevent people from seeing flaws in a plan (Puncochar & Fox, 2004). However, present research suggests that if the group did not create the plan, these effects might not exist. Participants using the Premortem Individual Method generated the same number of reasons as participants using the Premortem Method, and generated significantly more reasons than participants using the Worst-Case Scenario Method. The Worst-Case Scenario Method utilized a failure frame, but fixated participants on one specific outcome; having people list reasons why other outcomes may occur opens their mind to potential complications in the situation, and allows alternative possibilities to seem more plausible (Arkes, Guilmette, Faust, & Hart, 1988). This resulted in more reasons when the Premortem was utilized, regardless of the presence of a group. However, the group dynamic is an important factor in generating solutions and ways to strengthen a plan. Participants using the Premortem Individual Method generated significantly fewer solutions than participants using the Premortem Method and Worst-Case Scenario Method. This suggests that a Premortem can be conducted individually, but the plan will be stronger if individuals converge and discuss their reasons and solutions with others.

Implications

The underlying mechanisms of a failure frame and prospective hindsight are both necessary to generate more reasons for plan failure and ways to strengthen the plan. The validation of the Premortem in a Military field environment implies that this procedure is useful and can be easily implemented in MDMP or as a standalone plan evaluation technique for smaller teams. Moreover, Experiments 2 and 3 imply that the core mechanisms of the Premortem are different than simply imagining the worst-case scenario. Uncertainty reduction and a complete failure frame changes the way people approach plan evaluation, ultimately resulting in a more thorough analysis.

On a macro-level, Experiment 3 could support Meisser and Wulf's theoretical claims that the Strategy Scenario Method is superior to the Scenario Planning Method: instead of focusing on a best- and worst-case scenario as part of the plan evaluation process, implementing a Premortem can help teams develop more flexible strategies and resilient plans. The purpose of this thesis is to evaluate the Premortem Method as a lightweight plan evaluation technique; implementation of the Premortem into a larger plan evaluation tool is beyond the scope of the present research.

Limitations

It is difficult to do translational experiments in field studies and still control as many factors as one would like. While the results from the present series of experiments added valuable information to the limited field of Premortem research, there were several limitations in the present studies. The present experiments would have benefited from a

larger sample size; the results for fixation in Experiment 1 were statistically significant, but were based on 12 observers' subjective ratings. Furthermore, the difficulty of obstacles was unknown prior to execution. The obstacles were all different and were not systematically measured; some were extremely easy, and others were exceedingly difficult; this affected confidence and team performance. Since obstacles were categorized as ill-defined and well-defined problems post-hoc, they were not evenly distributed across plan evaluation methods.

In general, confidence was subjective; participants' ratings ranged from zero to 100 and the ratings were recorded only a few minutes apart. Future experiments measuring confidence would benefit from an explanation of the confidence scale prior to execution (Adams, 1957). Results from Experiment 2 suggests that an online platform is limited; in the present study, time was not controlled. Experiment 3 had a human facilitator; participants were timed while they wrote down their reasons and solutions.

Applying and Implementing the Premortem

The Premortem can easily be implemented with project groups and small teams; brainstorming literature suggests that smaller teams of about five are most effective, and individual brainstorming prior to group discussion is most effective (Bouchard & Hare, 1970). The Premortem is effective in helping small groups analyze their plans and generate ideas to strengthen the plan. Facilitation is important; Premortem participants must be instructed to imagine utter plan failure, brainstorm freely, and generate as many reasons for failure as possible. The facilitated group discussion ensures that all ideas are

discussed. Individual brainstorming of solutions and ways to strengthen the plan followed by facilitated group discussion allows participants to brainstorm freely, build off others' ideas, and generate more solutions than individuals can identify when working alone.

Future Directions

Military planning and decision making is a complex and varied field. Experiment 1 was conducted with ROTC Cadets during a leadership and team building exercise; while the Cadets were previously trained in Military planning techniques, their levels of experience varied and the situation was not exceedingly challenging or burdensome as one might imagine Military field environments to be. The Premortem effectively reduced errors and problem fixation in this type of field environment. Future research into application of the Premortem into Military planning and decision making would benefit from implementation during a more meaningful exercise where outcome is perceived as more important to the participants.

Within a more controlled environment, the next step in Premortem research is to examine the effects of groups and facilitators. In the present research, groups were pre-determined in Experiment 1 and ad-hoc in Experiment 3. A reasonable next study could determine the differences in confidence and number and types of reasons and solutions across both ad-hoc and pre-determined teams. The effect of facilitators should be examined as well; if teams are given only written instructions, will the Premortem still be effective? Previous literature suggests that nominal groups are as effective as groups

with highly trained facilitators, so there is reason to believe that a Premortem does not require human facilitation (Oxley, Dzindolet, & Paulus, 1996).

The end-state goal for the Premortem is a web-based or mobile application so a Premortem can be effectively conducted at an individual level on any type of plan. The Military has begun to integrate the Premortem into Commanders' toolkits; widespread knowledge of the Premortem and its underlying mechanisms as well as further development of this method as outlined above can help create a better tool for organizations and individuals. The present research contributes a large breadth of information to a previously limited field; this qualitative and quantitative analysis of the Premortem plan evaluation technique provides new insight into the importance of a failure frame and prospective hindsight mechanism as well as broadening the method's boundary conditions.

Conclusion

Plan evaluation techniques are often supplemented by anecdotal evidence and supporting literature, but not quantifiable evidence. Decision making and planning are complex domains that are enhanced by experience and expertise, but are challenging for novices due to an oversaturation of suggested processes and lengthy tools. The present research sought to quantitatively and qualitatively examine plan evaluation methods to provide support for the lightweight Premortem plan evaluation method, and develop recommendations for its implementation and practice in Military and boardroom environments.

References

Adams, J. K. (1957, September). A Confidence Scale Defined in Terms of Expected Percentages. *The American Journal of Psychology, 70*(3), 432-436.

Arkes, H. R., Guilmette, T. J., Faust, D., & Hart, K. (1988). Eliminating the Hindsight Bias. *Journal of Applied Psychology, 73*(2), 305-307.

Army Planning and Orders Production. (2005). Washington, D.C.: Headquarters, Dept. of the Army.

Bodwell, W., & Chermack, T. J. (2010). Organizational Ambidexterity: Integrating Deliberate and Emergent Strategy with Scenario Planning. *Technological Forecasting & Social Change, 77*, 193-202.

Brenner, L. A., & Koehler, D. J. (1996, March). Overconfidence in Probability and Frequency Judgments: A Critical Examination. *Organizational Behavior and Human Decision Process, 65*(3), 212-219.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "Planning Fallacy": Why People Underestimate Their Task Completion Times. *Journal of Personality and Social Psychology, 67*(3), 366-381.

Cojocar, W. J. (2011). Adaptive Leadership in the Military Decision Making Process. *Military Review*, 29-34.

De Dreu, C. K., & West, M. A. (2001). Minority Dissent and Team Innovation: The Importance of Participation in Decision Making. *Journal of Applied Psychology, 86*(6), 1191-1201.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. (2006, August 4). Frames, Biases, and Rational Decision-Making in the Human Brain. *Science, 313*(5787), 684-687.

Fischhoff, B. (1975). Hindsight/Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 288-299.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology: Human Perspectives and Performance, 3*(4), 552-564.

Griffin, D., & Tversky, A. (1992). The Weighing of Evidence and the Determinants of Confidence. *Cognitive Psychology, 24*, 411-435.

Grigsby, W. W., Gorman, S., Marr, J., McLamb, J., Stewart, M., & Schifferle, P. (2011). Integrated Planning: The Operations Process, Design, and the Military Decision Making Process. *Military Review*, 15-22.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased Judgments of Past Events After the Outcomes Are Known. *Psychological Bulletin, 107*(3), 311-327.

Klein, G. (2007, September). *Performing a Project Premortem.* Retrieved from Harvard Business Review: https://hbr.org/2007/09/performing-a-project-Premortem

Klein, G. (2008, June). Naturalistic Decision Making. *Human Factors, 50*(3), 456-460.

Klein, G., Moon, B., & Hoffman, R. (2006). Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems, 21*(5), 88-92.

Kobus, D., Proctor, S., & Holste, S. (2001). Effects of Experience and Uncertainty During Dynamic Decision Making. *International Journal of Industrial Ergonomics*, 275-290.

Koehler, D. J. (1991). Explanation, Imagination, and Confidence in Judgment. *Psychological Bulletin, 110*(3), 499-519.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980, March). Reasons for Confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118.

Lichtenstein, S., & Fischhoff, B. (1977). Do Those Who Know More Also Know More about How Much They Know? *Organizational Behavior and Human Performance, 20*, 159-183.

Meissner, P., & Wulf, T. (2015). The Development of Strategy Scenarios Based on Prospective Hindsight. *Journal of Strategy and Management, 8*(2), 176-190.

Mitchell, D. J., Russo, J. E., & Pennington, N. (1989). Back to the Future: Temporal Perspective in the Explanation of Events. *Journal of Behavioral Decision Making, 2*, 25-38.

Perez, C. (2011). A Practical Guide to Design: A Way to Think About It, and a Way to Do it. *Military Review*, 41-51.

Plous, S. (1995). A Comparison of Strategies for Reducing Interval Overconfidence in Group Judgments. *Journal of Applied Psychology, 80*(4), 443-454.

Pontis, S., & Blandford, A. (2016). Understanding "Influence": An Empirical Test of the Data-Frame Theory of Sensemaking. *Journal of the Association for Information Science and Technology, 67*(4), 841-858.

Puncochar, J. M., & Fox, P. W. (2004). Confidence in Individual and Group Decision Making: When "Two Heads" Are Worse Than One. *Journal of Educational Psychology, 96*(3), 582-591.

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences, 4*(2), 155-169.

Ross, K., Klein, G., Thunholm, P., Schmitt, J., & Baxter, H. (2004). The Recognition-Primed Decision Model. *Military Review*, 6-10.

Russo, J. E., & Shoemaker, P. J. (1992). Managing Overconfidence. *Sloan Management Review, 33*(2), 7-17.

Schmidt, J. B., Montoya-Weiss, M. M., & Massey, A. P. (2001). New Product Development Decision-Making Effectiveness: Comparing Individuals, Face-To-Face Teams, and Virtual Teams. *Decision Sciences, 32*(4), 575-600.

Shoemaker, P. J. (1993, March). Multiple Scenario Development: Its Conceptual and Behavioral Foundation. *Strategic Management Journal, 14*(3), 193-213.

Suler, J. (2004). The Online Distribution Effect. *CyberPsychology & Behavior, 7*(3), 321-326.

Thunholm, P. (2005, January). Planning Under Time Pressure: An Attempt Toward a Prescriptive Model of Military Tactical Decision Making. *How Experts Make Decisions*.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science, 211*(4481), 453-458.

Vane, M. A., & Toguchi, R. M. (2010). Achieving Excellence in Small-Unit Performance. *Military Review*, 73-81.

Veinott, B., Klein, G. A., & Wiggins, S. (2010). Evaluating the Effectiveness of the Premortem Technique on Plan Confidence. *7th International ISCRAM Conference.* Seattle.

Wright, G., & Goodwin, P. (2009). Decision Making and Planning Under Low Levels of Predictability: Enhancing the Scenario Method. *International Journal of Forecasting, 25*(4), 813-825.

Wright, G., Bradfield, R., & Cairns, G. (2013). Does the Intuitive Logics Method - And Its Recent Enhancements - Produce "Effective" Scenarios? . *Technological Forecasting & Social Change*, 611-642.

Appendix A

Squad: _____ Age: _____ Gender: _____

There are no right or wrong answers here, please report your opinion.

Using the scale below, rate the following events.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all **Neutral** **A lot**

1) By the last round, how easy was it for your squad to use the Premortem Method in this type of planning environment? _____

2) To what extent did your squad generate good reasons during the Premortem? _____

3) To what extent did using the Premortem change your squad's planning process? _____
   Expand briefly on how it changed (if applicable):

4) To what extent did using the Premortem improve solutions during the planning process?_____
   If less than 5, expand briefly on why:

5) Overall and relative to the baseline practices, to what extent did using the Premortem affect your squad's ability to recover from a bad idea? _____

   Please provide feedback about the integration with the task. If you were going to do the Premortem exercise again in this type of planning situation, what would you change, if anything?

Appendix B

You will be asked to evaluate the snow statue plan your group has developed for MTU's 2017 Winter Carnival. You will not be required to reveal details of your plan, what the final result will look like, or how the statue will be built. All information will be kept confidential, as outlined in the consent form.

Rate your confidence in your group's plan (0-100). A zero indicates that you have absolutely no confidence that your group will be able to execute this plan. A 100 indicates that you are completely sure that your group will achieve your goals.

Rate your knowledge of your group's plan (1-7). A one indicates that you are not familiar with the plan and could not explain it if asked. A seven indicates that you are very familiar with the plan and could teach it to others if asked.

Premortem Method:

Your group recently began building a snow statue for Winter Carnival. Imagine a few weeks into the future: Your snow statue is a complete fiasco. Your group has failed miserably. Take no less than 2 minutes to list as many reasons as you can as to why this failure happened.

Rate your confidence in your group's plan (0-100). A zero indicates that you have absolutely no confidence that your group will be able to execute this plan. A 100 indicates that you are completely sure that your group will achieve your goals.

Take no less than 2 minutes to list as many solutions for these reasons for failure as you can.

Scenario method:

Your group recently began building a snow statue for Winter Carnival. Imagine the worst-case scenario for your group's snow statue. Take no less than 2 minutes to list as many reasons as you can as to how this scenario might happen.

Rate your confidence in your group's plan (0-100). A zero indicates that you have absolutely no confidence that your group will be able to execute this plan. A 100 indicates that you are completely sure that your group will achieve your goals.

Take no less than 2 minutes to list as many solutions for these reasons as you can.

Rate your confidence in your group's plan (0-100). A zero indicates that you have absolutely no confidence that your group will be able to execute this plan. A 100 indicates that you are completely sure that your group will achieve your goals.

Rate your perception of your group's likelihood of success (1-7). A one indicates that you expect your group to not be successful in achieving your goals. A seven indicates that you are extremely assured of your group's ability to be successful.

Rate your personal level of commitment to the plan (1-7). A one indicates that you don't want to work on the statue and contribute to the success of the group. A seven indicates you are extremely devoted to seeing your group achieve its goals.

Appendix C

Participant Instructions/Scenario

You'll be participating in an exercise in which we'll gather your opinions about the university's Norovirus plans. Michigan Tech, like most other colleges and universities, has developed a plan for how to deal with a potential outbreak. Imagine that a planning committee has developed a plan for doing this and is seeking feedback about it. That's what we're going to do today—collect your feedback on their plan.

The information on the next page is the only information you'll receive about the plan and the facilitator doesn't have additional information, so do the best you can with the information you are given.

Do you have any questions at this time?

Please take a couple minutes to read the following plan that MTU could adopt for handling a disaster and provide your initial evaluation. There are no right or wrong answers here, we are interested in your opinion.

Example Scenario: Preparing for a Norovirus Epidemic at the University

Problem: MTU currently is implementing procedures to reduce the incidence of Norovirus, should an outbreak occur and an epidemic become possible. But what happens if an epidemic occurs and campus activities have to be curtailed or the campus has to be shut down? How will residential students be supported? How will the school year calendar be maintained? Because MTU has a mix of residential and non-residential students, it is much more vulnerable to such an epidemic than campuses that can be more easily quarantined.

Overview: The Lockdown Plan is intended as a last resort, in case of a serious epidemic at MTU and the Houghton area. The plan is to shut down the campus and, where possible, shift to distance learning to maintain the academic schedule.

Details: Phase I is to generate the announcement. The Lockdown Plan will be initiated at the decision of the President of the University, in consultation with key staff members. The announcement will be issued through the cell-phone advisory network and the MTU website along with other media (radio and social media). Ample warning will be given so that faculty, staff, and students can collect relevant papers and materials for working at home.

Phase II is to enforce the campus quarantine. Security forces will guard the access routes onto campus. Access to campus will be strictly restricted to only authorized personnel.

Phase III is to support residential students. They will be maintained on campus. Students who become ill will be re-located to special residential sectors and meals and medicine will be provided. Students will be allowed off-campus if they find it necessary to leave (e.g., doctor's appointment), but may be barred from returning. Special provisions will be made for students wishing to relocate off-campus and entering into their quarantine regimen.