TESTING RISK MANAGEMENT DECISION MAKING COMPETENCY OF PROJECT MANAGERS IN A CRISIS

Kenneth David Strang[1], Narasimha Rao Vajjhala[2]

[1] W3-RESEARCH, USA, ORCID: 0000-0002-4333-4399
Email: professor@kennethstrang.com
[2] AMERICAN UNIVERSITY OF NIGERIA, NIGERIA, ORCID: 0000-0002-8260-2392
Email: narasimharaonarasimha@gmail.com

Abstract: The objective of the current study was to use a rigorous controlled experiment simulating a project failure to measure how cognitive bias and competency impact a PM's risk management decision making in a crisis while controlling for other project and firm level variables including lying or faking responses. The MANOVA, repeated measures ANOVA controlled experiment and post-hoc analysis techniques were rigorous because the study took place in approximately the same point in time and each participant received all treatments. The 24 respondents in these repeated measures experiment outperforms most psychology factorial research design, which would require a 4 x 24 = 96 sample size to accommodate 3 treatments and a control group. We found bias significantly impacted PM risk management decision making in a crisis, but certification and competency resulted in the best decisions. Generalizations are cautioned due to the exploratory nature of this study. However, the literature review and methods were articulated well enough to encourage replications and extensions by other researchers.

Keywords: Project management, risk management, project manager, decision making, crisis, project failure, repeated measures ANOVA, experiment, MANOVA.

1. INTRODUCTION

Despite significant advances in information technology software and methodologies, projects have been failing at similar rates over at least the past 20 years (Pace, 2019). In fact, Ghossein et al. (2018) found the failure rate was 50% in European Union countries. This 50% failure rate was corroborated in the U.S. by the Standish Group (Kurek et al., 2017; Masticola, 2007). Other researchers found the longitudinal project failure rate in the U.S. ranged from 41% to 50% (Borbath et al., 2019; Eckerd & Snider, 2017). The problem is there was a lack of empirical literature explaining why about half of projects fail. The project manager (PM) is responsible for making important decisions to successfully manage the project, including escalating certain problems and risks to the sponsor. Not only was there a lack of literature on this topic, but the critical project success factors were often grounded in opinion survey data rather than upon factual evidence such as organization records or observation. Survey data is susceptible to biases of social desirability, false attribution, selective memory recall, and narcissistic motives (O'Boyle et al., 2019).

The objective of this study was to use a rigorous experiment simulating a project failure to measure how cognitive bias and competency impact a PM's risk management decision in a crisis making while controlling for other contextual elements including project and firm level characteristics. The research question (RQ) was how much do cognitive bias and other factors impact PM risk management decision making in a crisis. A crisis represents a severe risk event whereby the PM is accountable to make the best decision, whether that is to escalate to sponsors, a gating committee, or perhaps terminate the project if that authority level was granted in the charter. We argue that the ability of the PM to make good risk management decisions ought to positively impact project results albeit this will be affected by biases. We also argue that an experiment is necessary to collect objective data at a single point in time during a risk event rather than relying on memory recall and self-report perceptions. Prior literature determined certain PM individual level attributes likely impact project results, such as demographics, education, certification (Anthopoulos et al., 2016; Carlton & Peszynski, 2018; Damoah et al., 2018; Jennings et al., 2018; Pace, 2019), so these must be accounted for in an experiment, but researchers have not tested how bias impacts projects in a crisis, so new factors must be created.

The current paper is structured as follows. Section 2 reviews the relevant literature pertaining to empirical studies where project success or failure was the dependent variable. The key factors are identified and hypotheses are developed to answer the RQ. Since technology has evolved so quickly, and the coronavirus pandemic has impacted how PMs manage projects, more recent papers were favored. Section 3 explains the methods including the sample participants, ethics, and procedures. Section 4 lists the results and interprets the findings subject to their limitations. Section 5 summarizes the findings and discusses the implications as well as future study recommendations. The last section lists the references in scientific IEEE numbered format.

2. RELATED WORK

The literature review was focused on individual decision-making theories and factors, likely to be used by a PM, especially in risk management decision making during a crisis. This is hereafter referred to as risk management decision making. There was no rationale to re-review what we already knew about the key factors commonly tested in project success/failure studies, given that this was an experiment and we already knew approximately 50% of all projects had failed without robust empirical effect sizes. The only requirement in the experiment would be to establish a known context for all relevant factors beyond those of particular interest.

In other words, it would be necessary to develop an experimental scenario representing all individual, organization and global factors of a project, except for individual differences and variables being tested, to ensure the data were consistent for all participants. The different individual level PM attributes could be measured, such as age, gender, experience, education, certification, and competency. For this reason, the literature review concentrated on risk management decision making, at the individual level of analysis. Below the studies with project success or failure result as dependent variables are reviewed to identify the relevant independent factors along with the applicable methods and findings.

2.1 Common Predictors of Project Results

There were four studies of project result as a dependent variable, which were selected from the literature review search short list because they were most relevant to the current study. The three key reasons were: Each had very large sample sizes, each was empirical, and the research design with methods was explained in detail. Those studies are reviewed below in order of largest sample size first. Although each study measured project result in slightly different ways, the common finding was that one or two quantitative data types were posited to indicate success as being on time, within documented scope, having acceptable quality and under or at budgeted cost. In all four studies more than one dependent variable was used to capture each of the above project results separately as scope, time, cost, quality, and or overall success. Sometimes the dependent variable was documented as a nominal or binary field with a yes (1) or no (0) value.

The first paper reviewed was the rigorous monolithic quantitative study by Ghossein et al. (2018). This was an exploratory empirical study of the factors impacting the high project failure rate of 50% in Europe. They applied Pearson correlation to quantitatively examine mixed data type variables from 59,816 public procurement projects of European Union (EU)-based member nations to determine which factors were related to project results. They declared a 90% confidence level, which is acceptable for exploratory studies. Most of their factors and dependent variables were binomial (1 = success or 0 = failure). They sampled small-to-medium-sized-enterprises and large firms. They examined the relationship between firm size, structure, certification/training, experience (in years), exporter status, foreign ownership, access to finance, crime losses, growth rate per GDP capita, level of development, land area owned, service sector, geographic region and procurement project result. This was secondary data obtained from the EU. Their findings indicated that in terms of innovation, ISO quality had a far greater effect on manufacturing firms than services firms. They claimed effective project management systems were negatively correlated with corruption faced by EU businesses. However, they did not clearly define how the corruption factor was defined.

Despite the monolithic sample size and several significant coefficients, the Ghossein et al. (2018) study contained a few limitations of missing hypotheses, small effect sizes, and lack of proven causality. This could be improved in future studies by using a predictive regression model design. We could also add that the logarithmic transformations could obscure factor interaction and the lack of proven causality was likely due to a correlational design. However, it can be asserted their design was rigorous, the sample large, and they observed several statistically significant effect sizes for associations between variables of interest in the current study. It was also one of the few empirical studies examining project success in Europe, beyond single case studies, with articulated research design methods encouraging replications. They found the age and size of the organization were associated with project results yet the effect sizes were negligible. Revenue was not found to be related to anything. They found no statistically significant relationships between firm size, structure, certification/training, experience in years, exporter status, foreign ownership, access to finance, crime losses, growth rate per GDP capita, level of development, land area owned, service sector, geographic region and procurement project result. Interestingly, Ghossein et al. (2018) found that ISO quality status had a positive influence on product innovation, process innovation and R&D spending for manufacturing firms, but no statistically significant effect for service firms.

Borbath et al. (2019) published a large rigorous quantitative study of high-priced defense industry project failures with budgets over $1M USD. Their sample size was large at 14,836 projects across the three U.S. Army, Navy and Air Force divisions. They collected secondary data containing project metrics from U.S. government contracts. They used Spearman correlation to examine if contractor project result was related year-on-year across three ordinal variables: Cost, schedule, and technical quality score. Their findings were disappointing in as far as they replicated the *a priori* literature to find only that past performance was an indicator of future performance. Nonetheless, they replicated many individual PM factors which impacted project results, as already identified in the above literature review. They argued PM age was likely to be high when experience was high. Additionally, they asserted PM certification was a key predictor of project result which was corroborated by Eckerd and Snider (2017) as well as well as by Barrows et al. (2020).

The next large rigorous quantitative study reviewed was a benchmark paper in the project management discipline by Eckerd and Snider [6]. They examined which factors caused defense industry projects to succeed or fail. Theirs was one of the few published articles where project manager certification, education, and other common factors such as experience, were tested with a large sample size to determine if project result could be predicted. They used nominal or ordinal data types for many factors and interval data types for dependent variables. They applied generalized least squares (GLS) regression to quantitatively analyze procurement data from 1,073 large U.S.-based defense projects in the Navy, Air Force and Army divisions completed during 1997-2010 with the goal to empirically identify what caused projects to succeed or fail. There were two interval data type dependent variables at the project level: Cost variance and breach.

Eckerd and Snider (2017) explained that cost variance was the difference between the final project costs minus the original baseline estimate, which can be interpreted as: A cost variance higher than zero was considered a failure. A breach indicated a sufficiently large deviation away from the required procurement project schedule, cost or quality, which also indicated a failure. Scope variance was not captured as a project result because government defense procurement projects have a specific statement of work which contractors bid on, and this can be modified only through approved change orders. They could not find any significant relationship between project leader attributes such as age, gender and experience regressed on the dependent variables.

Although Eckerd and Snider (2017) produced very interesting regression estimates, they did not they did not report customary sample descriptive statistics (instead their estimates referred to military project characteristics), variance inflation factor estimates or partial correlations. The GLS regression odds ratios (OR) and beta coefficients (B) indicated four U.S. government defense industry project categories were significantly more likely to have poor cost variance or breach probability. Specifically, these were Aircrafts (OR=7.38, B=-.078, p<.05), ground technology (OR=10.46, B=-.05, p<.05), ships (OR=9.52, B=-.092, p<.05), and space systems (OR=13.40, B=-.095, p<.05). To interpret their GLS regression model OR effect sizes, space system projects were 13 times more likely to breach cost, schedule or quality requirements although the beta coefficient of -0.095 was a relatively small negative budget deficiency in the regression formula. The key limitation of their findings was that although the coefficient sizes such as OR were strong, the departmental division factors were at too high a level of analysis to be useful or generalize outside the military government. Also, they were missing model effect sizes beyond a Wald estimate. A Wald estimate is similar to a chi square goodness of fit between a null and final model, but it does not measure effect. Nonetheless, their study serves as a valuable template for researchers to extend project management analysis in the defense industry.

Finally, the post-pandemic empirical study by Brandon et al. (2022) was a comprehensive quantitative intensive exemplar which could motivate other research in the project management discipline. They used ordinary least squares (OLS) regression and step wise regression (SWR) with SPSS to examine which individual level PM factors predicted overall expertise and project success. They surveyed PMs using an outside agency, and after cleaning their data, the final sample size was 625 respondents. There were many notable features of their study. First they clearly described the problem rationale for their study and grounded it in relevant literature. This included an explanation of potential significance. Second, they articulated their methods in extensive detail, with comprehensive step by step result explanations, facilitating replication and understanding by emerging scholars.

Interestingly, Brandon et al. (2022) leveraged the LinkedIn professional social media application as part of the sampling protocol. They examined the LinkedIn project manager community groups as a population source. LinkedIn was also the medium used for soliciting population sample respondents in the current study. They described their sample as random but self-selected, with a required minimum based on a guideline of 15 participants per 36 exploratory factors which would result in 540 not the 500 they stated. Later they revealed the G*power tool was used to calculate the required sample size — albeit redundant, it was useful to know the power was 100% and the sample size of 500 would be adequate (they attracted 679 respondents so this was exceeded).

Fourthly, Brandon et al. (2022) were one of the few researchers to list the effect sizes in their results. They found four factors significantly predicted overall expertise in their step wise linear regression model, with $F(4,620) = 10.874$, $p < .0005$, adj. $r^2 = .066$. The four factors were: Age as first PM, number of months managing projects [experience], months per project, and education level. They then repeated the linear regression with the second dependent variable project success, identifying two significant factors: Age as first PM (beta = 0.168), and number of months managing projects (beta = 0.066), with r^2=0.025 (p. 28). In their OLS model, they claimed four factors predicted expertise: Age as first PM, months managing projects, projects per month, and education level. However, when examining their ordinal least squares regression estimates, only education contributed any meaningful statistical impact on the expertise dependent variable with OR of 1.257, which they referred to as “Exp(B)” (p. 29) – all other factors were near or less than a 1 OR, with model effect size of $r^2 = 0.068$. Only their second stepwise regression model was relevant to the current study because the dependent variable was percent success. In that model, they found two factors predicted percent success: age as first PM ($B = 0.013$) and months managing projects ($B = 0.088$), with a weak effect size of 2.5% ($r^2 = 0.025$). The months managing projects, although a small beta of +0.09 (rounded), indicated PMs with more experience achieved more success, noting the independent factors and dependent variable was self-reported. Likewise, the smaller beta of +0.01 (rounded) for age as first PM can be interpreted as younger PMs tend to have higher percent success, which does not make theoretical sense. They were two negative betas in the stepwise linear regression predicting expertise: Age as first PM ($B = -0.084$) and months per project ($B = -0.163$), with a model effect size of 6%. This contrasted with the positive beta of +0.01 for age as first PM to predict percent success. This could be interpreted as older PMs perceived they had more expertise but less percent success, which was puzzling. Nevertheless, age and years of experience clearly need to be included in the current study as factors. PM education is relevant to consider as it had the highest OR, although it predicted expertise but not percent success.

There were a few elements in the well-crafted study by Brandon et al. (2022) which could be improved for future replications. First the most frustrating aspect of their study was trying to read the 3 across page format! A single page display would be much easier for scholars to navigate. Second, their research design started as correlation with RQ1 "What is the relationship of the six experience factors and Overall Expertise" and RQ2 "What is the relationship of the six experience factors and Percent Success?" (Brandon et al., 2022). The hypotheses were relational tests but this transformed to a predictive model in the results: "A stepwise regression was run to predict Overall Expertise from six factors based on self-reported experience factor" (Brandon et al., 2022).

Brandon et al. (2022) sometimes wrote in difficult to understand nomenclature. The iterative data cleaning was verbose, decreasing the sample size several times from 679 down to 625 as they proceeded through outlier and residual analysis over several pages. The data cleaning was important to mention but the excessive details could have been left out with no adverse implications. More than one anthropomorphism error was noticed – this occurred when they attributed their actions or findings to external researchers, for example "data were analyzed for reliability and for statistical test assumption compliance for stepwise linear regression (Laerd, 2015a) and ordinal logistic regression (Laerd,2015b)" (Brandon et al., 2022). Anthropomorphism errors could have been easily avoided by not citing a source when describing what was done within their study. A fourth concern was most of the literature review did not seem related to their independent factors (e.g., risk management, self-efficacy) or dependent variables, unless there was another theoretical perspective missed in our review.

Brandon et al. (2022) stated 36 factors were available but only 6 indicators were mentioned. Projects per month, months per project, months managing projects, education, age when first PM, and expertise were the 6 variables. This raises the fifth point being there was no explanation how the 2 dependent variables were formulated (e.g., data types, value ranges). It was unclear how did they captured or calculated the 'percent success' dependent variable. Also, there were no descriptive statistics showing either the sample demographics or estimates of the factors and dependent variables. The results cannot be reliably generalized without sample descriptive estimates. A sixth design issue to consider in the future was that some of the independent factors were more likely dependent variables referring to the same point in time as project results, or were inversely interrelated, and would therefore create variance inflation, namely projects per month and months per project. It did not seem rational to capture age when first PM but perhaps this was done to improve objectivity instead of asking for years of experience. This relates to the seventh idea, that instead of within group analysis using multiple OLS and stepwise regression models separately on a single dependent variable, a multivariate MANOVA or Friedman between groups technique could be applied to compare project results dependent variables across the independent factor levels, which would incorporate all estimates and standard error in one model.

Finally, Brandon et al. (2022) claimed the correlation effect size of "Overall Expertise (R2 = 0.874 and 0.963) … were very weak effect sizes" (Brandon et al., 2022). Those were not weak effect sizes. However, for the expertise dependent variable, the effect size of the stepwise linear regression model was 6% and 7% for the ordinal linear regression (p. 28). Those large effect sizes of 87% and 95% would be statistically impossible in any of their regression models. Did this refer to an unreported 36 factor model or average bivariate correlations of the 6 reported factors? Nonetheless, their paper provides an excellent PM research design model for new scholars if the above points are properly addressed.

2.2 Common Factor Hypotheses Formulation

Based on *a priori* literature (Barrows et al., 2020; Borbath et al., 2019; Brandon et al., 2022; Eckerd & Snider, 2017; Ghossein et al., 2018; Hughes et al., 2017; Pace, 2019), it became clear that certain individual level PM factors were common place predictors of project results across all empirical studies. Only one study specifically tested expertise or competency as a project result (Brandon et al., 2022), while most other researchers integrated competency into independent factors such as certification or experience (Barrows et al., 2020; Borbath et al., 2019; Eckerd & Snider, 2017; Pace, 2019). Certification was posited to be related to project result in several studies (Barrows et al., 2020; Borbath et al., 2019; Eckerd & Snider, 2017; Pace, 2019). Given the common testing of competency and certification to predict project result, it makes sense to capture those two factors in the current study, if only to remain comparable to extant literature. Experience and certification are easily captured as a self-reported PM attributes, but competency may be difficult to objectively capture as a perception. The consensus from the recent literature identifying the most important PM individual level of analysis to predict project result were: PM age, gender, years of experience, education, competency (expertise) and certification. The following positive correlation hypotheses were developed to answer some aspects of the RQ.

H1a: PM age is related to good project results;
H1b: PM gender is related to good project results;
H1c: PM certification is related to good project results;
H1d: PM education is related to good project results;
H1e: PM experience is related to good project results;
H1f: PM competency is related to good project results.

2.3 Risk Management Decision Making in Projects

The above literature review identified potential individual level PM attributes including demographics, education, experience, and certification, which were thought to impact project result. Organizational level factors may also impact project result, but the scope of the current study was on the individual level of analysis. However, in a controlled experiment, the organizational level and other contextual factors must be accounted for in some manner, which will be addressed in a later section. An important individual level factor of interest for the current study was how PM risk management decision making competency impacts project result.

The concept of risk management decision making in projects was introduced by Doug Barlow in 1963 (McShane, 2018; Raftery, 1994). His version adopted the economic cost of risk theory from the finance discipline while adding concepts for risk avoidance, risk mitigation, and risk retention (McShane, 2018). Risks were classified into four major categories, namely, hazard, financial, operational, and strategic risks (Callahan & Soileau, 2017). A crucial point in mitigation after a risk event was achieving the effective balance between risk retention (continuing projects such as through approved change orders) as contrasted with risk avoidance (cancelling projects or transferring responsibility to subcontractors or insurance companies).

The Committee of Sponsoring Organizations of the Treadway Commission in 2004 CSOTC (2017) documented new perspectives for risk management decision making. CSOTC was an accounting-oriented consortium who emphasized strategy and performance variables with a strong emphasis on organizational culture including risk avoidance (Prewett & Terry, 2018) and reputation management (Bohnert et al., 2019). Interestingly, risk avoidance and reputation management refer to individual behavior in psychology (Conner et al., 2013), well beyond accounting nomenclature. Other variables in their framework included downside side risks associated with disaster planning insurance and hedging activities while upside risks were focused on business strategy adjustment and reserve capital development (Callahan & Soileau, 2017). The project management discipline borrowed these theories (Goodwin & Strang, 2012) and many researchers have integrated these concepts into other professions or topics (Korstanje et al., 2020). The key initial point of failure is at the individual PM risk management decision making stage since the PM must first formally identify a risk and then make the critical decision or a recommendation that a project ought to be continued or terminated. This warrants a deeper look into the individual risk management decision making theories and models.

There are individual cognitive and psychological dimensions at play during risk management decision making such as bias (Conner et al., 2013). Although individual level risk models have been applied in several contexts, there was limited literature on examining the cognitive dimensions of PM risk management decision making in the context of a project failure. It is generally accepted in psychology that complex decision making, such as terminating a failing project, cannot be taught solely during training or certification, nor can these competencies be articulated in organizational procedures due to the required tacit knowledge such as perceptual skills, pattern repertoires, rich mental models, and sense-making (Klein et al., 2016). Therefore, bias may be a cognitive predictor of project failure. It is difficult to measure cognitive decision making in a failed project because PM bias is tacit and memories fade when accessed beyond the point of failure. Another approach would be needed to determine if hidden PM decision making factors like bias could be the reason half of all projects fail.

The individual level decision making theories and models could be grouped into three major themes, each with many theories or models. Cognitive models, the largest group, describe the mental or emotional factors and processes people use to make decisions. Most of these originate in the psychology discipline and there are numerous replications with authors putting their own terms on models which are theoretically unchanged from years ago. Popular cognitive models studied in management contexts have included bounded rationality, rational (economic) model, incremental (trial and error), garbage can model, prospect theory, and recognition-primed decision theory. The second group, attribute models, could be used to describe the family of individual decision-making concepts where physical traits, skills, competencies, knowledge, and or personality describe PM decision making. This may include individual differences, leadership theories, experience, education, culture and so on.

The third group, structured or enterprise models, cover organizational operating procedures or other deliberate processes, requiring the PM to execute specific steps and then evaluate the data to make a decision which may call for escalation to a higher authority to determine the ultimate course of action (Elbanna & Child, 2007). These theories include the political view model, program model, Lean Sigma/DMAIC quality assurance/quality control, agency theory and the garbage can model. The multiple perspectives concept seems to fit this third category, by combining ideas from both of the above, but in an organizational context, such as leadership attributes and the rational model of decision making. All models implicitly or explicitly include bias discussed above, but the advantage of models is they are more operational for instrumentation in experiments since the factor relationships are a priori, they can be measured. In the current study we are focused on cognitive bias aspects of the PM because we are using an experiment to control for the organizational factors found in structured model and we are measuring the relevant physical level attributes. Therefore, we can concentrate on cognitive bias concepts of risk management decision making.

The cognitive bias perspective of individual decision making refers to emotional or thinking processes for solving complex problems in projects beyond ordinary business as usual activities or standard operating procedures. For example, deciding what to do when an experienced team member dies or a supplier fails to deliver critical resources before a major milestone. A natural or man-made disaster or similar crisis would constitute a relevant complex risk management decision making scenario. By contrast, underperforming resources or scope creep could be classified as business-as-usual activities for an experienced PM.

Effective decision making during stressful risk events like a crisis or disaster requires tacit cognitive competencies which cannot easily be learned such as perceptual skills, pattern repertoires, rich mental models, and sense-making skills (Klein et al., 2016). On the other hand, Klein et al. (2016) argued that some decision-making skills can be taught, such as adopting a macro-cognitive problem-solving mindset, recognizing subtle cues of boundary criteria in a crisis to trigger abandonment of an inevitable failure. A critical finding was that bias affects decision making, such as heuristics, herding, and prospects, which could result in irrational choices even when all relevant information is available (Goyal et al., 2021).

Heuristic bias refers to the reliance on mental shortcuts in decision-making. These heuristics include overconfidence, representativeness, and anchoring (Goyal et al., 2021). Overconfidence is considered the most significant and dominant of all the biases even when others are present (Goyal et al., 2021). Overconfidence refers to the tendency of the decision-makers to believe that they are better than what they are, or that their decision will be the best (Goyal et al., 2021). Some of the factors contributing to overconfidence include the individuals' age and organization variables (Goyal et al., 2021).

Unrealistic optimism bias means a PM tends to overestimate the positivism of the future while representativeness refers to the overreliance on stereotypes influencing their decision-making (Jordão et al., 2020). For instance, some investors may consider the past returns of a firm to represent the future performance of a firm, which can lead to investors making erroneous decisions based on a heuristic that is not essentially valid in an investment scenario. Optimism bias is responsible for failures in setting deadlines and in cost-benefit analysis (Goyal et al., 2021). The anchoring bias is the tendency for decision makers' judgments to be influenced by the opinions or initial information (Jordão et al., 2020). Anchoring relies on reference points created by decision-makers to assist them in the decision-making process including ego or personal gain (Goyal et al., 2021). Age and gender were among the demographic factors that impacted the anchoring heuristic, with women and older age groups tending to have a higher degree of reliance on anchors (Goyal et al., 2021). Therefore, it is wise to capture demographic factors including gender, age and experience for control in an experiment because if they were related to project failure then bias could be considered redundant in the model (due to physical PM attributes predicting project failure).

Prospect theory (PT) was one of the most significant models developed for risk decision-making and it leverages many cognitive bias concepts discussed above which can be assessed in experiments (Kahneman & Tversky, 1979; Tapas & Pillai, 2021). PT leverages loss aversion theory, whereby investors tend to be risk-averse when it comes to gains but risk-seeking when presented with potential losses (Tapas & Pillai, 2021). In this theory, risk decision-making is viewed as a choice between value prospects or gambles based on risk-seeking or risk-averse behavior informed by loss aversion, reference points, and decision weights (Kahneman & Tversky, 1979). Tversky and Kahneman (1992) proposed cumulative prospect theory (CPT), extending the expected utility model of PT (EPT) by applying the cumulative function separately to gains and losses. In the extended PT model, individuals may claim to be logical decision-makers, but they are susceptible to cognitive biases and overlook logical choices - when presented with a choice, an investor tends to choose the one presented in terms of potential gains (Kahneman & Tversky, 1979). Some elements of EPT could be ideal for explaining individual risk management decision making in troubled projects because one of the factors in the framing process is clearly experience, while the prospect valuation process likely involves personal bias. The personal bias could be that a PM will want to maintain a good reputation, earn more income, recognition, social acceptance, and so on.

2.4 Risk Management Hypotheses Formulation

In the current experiment we argue it is reasonable that a competent and certified PM would cancel a failing project. We also posit competence and certification could go hand in hand meaning both would be present in successful PMs. Therefore, we believe the reason half of all projects are failing is due to bias impacting the PM risk management decision making process. It is argued personal bias could impact the PM's decision more than merely technical competence, even to the extent that a doomed project may be continued solely due to personal bias. The concepts of value prospects along with ego/overconfidence bias are argued to be the most plausible hidden factors impacting an experienced, competent PM. The is because a PM is chosen due to previous project successes, which means they would have realized economic gain as well as recognition, so it is argued a PM will be influenced by emotional ego and economic gain prospects when making a complex decision. It is less plausible that a PM would be risk avoidant since most projects have inherent internal risk and unknown external risks. Additionally, risk management is a PM function intended to manage risks. Thus, it is posited that PT and EPT would be unlikely to function as a model, although some of their underlying bias concepts would likely prevail. While it may be possible other bias would influence a risk management decision making, it is argued they would be over shadowed, subsumed or could be measured by economic prospect bias indicators.

Finally, the bias factor must be operationalized. Since there were no published experiments examining bias in successful or failed projects, we can consider how bias has been tested in psychology. Most often bias was presented as a monetary gain or loss (Kahneman & Tversky, 1979). We could reasonably assume a gain would encompass economic and emotional gains while a loss would result in economic and emotional losses. Thus, we could create a rudimentary composite factor representing bias as an economic incentive with project success implications compared to a lack of bias as no economic incentive without implied project success. The PI could not explicitly state failure as a condition or consequence in the experiment because that would have created a negative bias since it is unlikely any experienced PM would willingly associate with a failed project even in a study. Thus, bias must be captured inherently in the treatment. A placebo or no bias condition would also be needed as a comparison. In order to answer the RQ, the implied project result is whether to cancel a failing project during a crisis. The following comparative hypotheses, contrasting certified versus not certified PMs, and competent versus not competent PMs, were created. These will be operationalized in the methods section since the outcome variable will be measured in the experiment.

H2a: Certified PM's are more likely to cancel a failing project without bias (basic decision);

H2b: Competent PM's are more likely to cancel a failing project without bias (basic decision).

H2c: Certified PM's are more likely to cancel a failing project with bias (biased decision);

H2d: Competent PM's are more likely to cancel a failing project with bias (biased decision).

3. METHODS

3.1 Research Design and Analytical Approach

A post-positive ideology was applied by the researchers because the RQ was deductive, data-driven and predictive in nature. The confidence level was set at 95%. Since this was an exploratory experiment, an 80% power level was desired. Means, standard deviations (SD) and ranges were calculated for continuous data types. Frequencies, percentages, ranges and median or mode (if needed) were reported for the other data types. Nonparametric correlations as well as repeated measures tests with post-hoc comparisons were used to evaluate the hypotheses. Normality tests were conducted for each technique as needed. Repeated measures ANOVA and MANOVA with post-hoc tests were applied, as elaborated upon below.

A pragmatic ideology was also applied to the literature review, using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) technique. The authors applied PRISMA by continually eliminating articles in the search results if they were not empirical and related to project results/performance. Excel was used to weight each article and sort the higher priority papers, to provide the final short list for discussion in the current paper. Many papers were eliminated due to fatally flawed research designs, failure to report effect sizes, or other serious impediments, namely the project management association journals. For statistical analysis, R-Studio, Dataiku DSS, SPSS, and Tableau software were utilized.

This was a repeated measures experimental design because all participants were given all treatments, the competency test and both experimental conditions. A repeated measures-controlled experiment at a single point in time is considered one of the strongest scientific empirical research designs to test human behavior. The reason a repeated measures-controlled experiment is robust is that any variable not observed or captured in some manner is controlled by making it static and known to participants. Additionally, every participant received all treatments in the same sequence with identical controls. Individual differences remain the only varying factors with everything else controlled, thereby eliminating isolated subgroup interaction between specific conditions and individuals. However, the current study was limited to a quasi-experiment since it was conducted asynchronously online (not in a lab) yet it was confidential, the duration was one week, participants were randomly selected and they had no contact with one another. It was also possible that lingering sentiments after the first treatment may have cognitively impacted the second treatment, since it was a repeated measures design by purpose.

All *a priori* global and organizational level factors were controlled including firm size, revenue, industry type, quality registration, process maturity, standard operating procedures, project/team context and the risk events. The independent factors were captured, namely: PM certification, PM education level, PM age, PM gender, and PM years of experience. PM competency was a continuous ratio data type calculated from a test. The dependent variables formed in the experiment represented the decision to cancel a failing project, a ratio data type (1 to 5), with higher values meaning a good decision - to cancel. A sliding scale was used to capture whole numbers with decimals. The treatments consisted of a basic risk management decision with no incentives while the biased risk management decision was configured with an incentive to capture ulterior motive bias.

3.2 Ethics, Participants and Data Collection

The first author was the principal investigator (PI). The PI was a licensed practicing PM in a global multinational company, a part-time visiting professor at different universities, a certified research professional, and approved by the Collaborative Institutional Training Initiative for conducting social science research with human participants. The PI conducted the initial literature review, and wrote the first paper draft. The PI designed the experiment and received approval form the internal research control board (ICRB) at his employer noting there was low risk anticipated as the sample was not a protected group and there were no adverse outcomes anticipated. The PI executed the experiment online after ICRB approval. The secondary author conducted an extensive secondary literature review, and collaborated with revising the paper. The second author was also a licensed PM and a professor of computer science. He presented the research design and a preliminary paper to peers for constructive feedback at the 15th IADIS International Conference of Information Systems, March 12-14, in Lisbon Portugal. The authors reported no conflicts of interest. No additional external funding was received for this study.

Initially, the PI randomly identified candidates for the experiment from the global population of approximately 23,000 professional-level PMs in a relevant LinkedIn group citing PM in their profile. This was believed to be the global population of interest but from a brief demographic analysis admittedly it contained mostly U.S.-based PMs. In fact, all respondents were based in the U.S., which somewhat limited the generalization. A sample size was not as critical an issue in a controlled repeated measure experiment as long as at least the sub group size was 5 or higher (which it was). In a repeated measures experiment, all participants receive every treatment. Since PM experience and certification were proven to impact project result, a minimum of 5 years of experience was used as a purposive sampling criterion and an attempt was made to include roughly a balance of certified versus uncertified PM's in the sample. Subsequently, the PI developed a short list of 16 qualified participants, which was expanded to 24 after a second round of solicitations. Participants were invited using a private email describing the brief online experiment as a safe beneficial research project with a token payment up to $40 depending on answers to the experimental condition questions. A few brief screen questions were asked to apply the sampling criterion. Thus, the final sample size was 24 PMs.

3.3 Measures and Procedures

The experiment was created as a project brief document describing a large military software development initiative, with exactly 50% of the anticipated 12-month duration completed. Participants were given ethical disclosure and consented. They were told they would receive a minimum of $10 and up to $40 depending on performance in the experiment. Actually, after the experiment, all participants were given the same $40 since the incentive scale was an illusion designed into the treatment. A project dashboard was displayed with metrics representing all the organizational and global factors to ensure there were no missing facts and that the contextual data would be identical for every participant. There was a statement highlighted in bolded font that any delays beyond 30 days must be presented to the gating committee for a project go/no-go decision or instead the PM may issue a cancellation notice. Below the dashboard, participants were informed a risk event had just been detected and they the PM needed to make a decision. Participants were told the coronavirus COVID-19 pandemic was now impacting the team because apparently someone had tested positive.

This first activity was a competency test to determine how accurately the PM could calculate the risk impact on the project. The risk event was that 1 of 2 senior programmers who were to create the online security frontend (not yet started) suddenly died due to COVID-19. The security program had the projected 5-week 30-day duration (60 effort days of 2 people) and was on the critical path to start now. There were no similar resources currently available on the team so now only 1 programmer was available. Participants were given a short open-book style test, allowing 30 minutes to complete it, to determine their competency in calculating standard deviation risk on this revised activity. They were told it was permissible to use books, notes or programs if needed but to do the work themself. Participants were told to help mitigate this COVID-19 risk event by providing a new estimate for programming that module by using the PERT technique with standard deviation to quantify this risk. They were given the estimates from a focus group of experienced team members: Their most likely effort for 1 programmer operating in isolation was 30 person days, an optimistic time was 12 days but the pessimistic value was 60 days. Participants were asked for the expected effort in days under this risk and allowed 10 tries to get the correct answer of 32; the maximum score was 10, reduced by 1 point for each invalid answer. Participants were not given the answer or their accuracy - to reduce bias. In addition to the ratio data type, a binary competency field was created as 0<60% or 1 for pass.

The first treatment was a no-incentive condition to see if the participant would terminate the project when they ought to due to the severity of the risk and according to the mandate above. In order to make it clear that a severe risk event had occurred which ought to stop the project, participants were told that suddenly the second programmer also died of COVID-19. Participants were told to save time the system automatically recalculated all project activity effort estimates. Participants were shown the project dashboard with the bolded estimate of an additional 64 days effort would be needed. The dashboard contained the same bolded statement as before that any delays beyond 30 days must be presented to the gating committee for a go/no-go decision. Participants were told they could keep the project going, apply their competency by creatively working with the team to find ways to meet the original schedule with less than a 30-day delay – OR – they could raise a project cancellation notice to the gating committee and wait on the bench for a new project to be assigned. Participants were told there were more questions but first they were asked their likelihood to issue the cancellation notice. They were given a sliding 1-5 ratio scale with a decimal, using these prompt headings: 1=strongly disagree (don't cancel), 2=disagree (don't cancel), 3=undecided, 4=agree cancel, 5=strongly agree cancel. Any continuous number between 1 and 5 would be stored according to the respondent's cursor movement on a visual sliding scale. Respondents were given 15 minutes to complete this exercise.

The second treatment was an incentive condition. Participants were told to consider their previous answer void (never happened). Participants were asked to reconsider mitigating the risk event above, due to new information becoming available. They were told one of the marketing directors, a key stakeholder in a different division, offered to increase the PM salary by a 400% bonus if they did not cancel the project. The marketing director suggested the PM could fast-track the hiring of several cheap but talented computer science students under the radar to program the security model in the original budget without having to delay the project beyond 30 days.

To make the experiment feel realistic, participants were deceptively told they would get $40 instead of the $10 in this experiment if they did not cancel. Now, with a 400% incentive, participants were asked the same question as above with the same ordinal scales. They were given 15 minutes to complete this question. Note participants were actually given the full reward at the end of the experiment so everyone received the same payment regardless of their performance. Participants were then asked for their demographic data to capture important individual PM factors. Age and years of experience were stored in the database as ratio data types. Gender was indicated as 0=female, 1=male. The highest education level was recorded as an interval data type using these scales: 1=school only, 2=associate, 3=bachelor, 4=master, 5=doctorate. PM certification was created as a binary field, 1=yes, 0=no.

The last experimental treatment was arranged to force the PM to choose between a personal value prospect gain including implied reputation/emotional gain, versus no gain, following documented risk management procedures of issuing the cancellation notice since the project would clearly be delayed more than 30 days. It was understood in the PM profession that cancelling a project is not a desirable outcome, even if there would be no value prospect loss. After the experiment, each participant was debriefed by the PI, to arrange payment of the full amount regardless of their experimental performance, to ensure they were satisfied, as well as to document their perceptions and insights.

4. RESULTS AND DISCUSSION

4.1 Preliminary Analysis of Experimental Data

There was no experimental sample mortality or respondent attrition since all 24 participants completed the experiment confidentially and answered all questions in the same timeframe of 1 week without knowledge of one another. There was no data cleaning needed but the experiment was controlled – only allowable ranges of responses were accepted. However, the PM demographics were checked for outliers, such as age, experience, certification, education level, and gender with no outliers. For example, an age below 20, only high school education level, or years of experience above 40, would be suspicious for a PM in this sample. Note also that social desirability and other survey response faking red flag questions were not asked because this was a controlled experiment with an incentive ─ motivation was inherent due to the incentives and lying would actually be an acceptable response outcome to theoretically indicate a poor risk management decision.

The important descriptive statistics including means (M), standard deviation (SD) and Spearman correlations are listed in table 1. The mean age was 42.5 (*SD* = 8.3), ranging from 29 to 59. Most respondents (71%) were male. The average years of experience was 11 (*SD* = 4.3), which ranged from 5 to 21. All participants had at least a bachelor degree and this was the most common (minimum = 3, median = 3 and mode = 3), while a few had a master and one had a doctorate. Approximately half (50%) claimed to be certified. Slightly less than half of participants passed the competency test with a score of 60 or higher (*M* = 55%, *SD* = 31%). This was roughly the same as the global project failure rate cited earlier.

Shapiro-Wilk (SW) tests were conducted to check for multivariate normality of the measured variables. The results indicated competency was normally distributed with SW = 0.92 (p = .058), but the dependent variables were not: Basic decision (SW = 0.784, p < .001), and biased decision (SW = 0.882, p = .009). The SW results indicate caution must be observed when using certain statistical techniques with the dependent variables. Competency was transformed to facilitate ANOVA and MANOVA group comparisons. A new variable competent was calculated to indicate a 1 if the respondent had a score of 60 or higher and 0 otherwise.

4.2 First Set of Hypothesis Test Results

Correlation tests were used to test most of the hypotheses. Additionally, since the variable types ranged from categorical to ratio, the nonparametric Spearman correlation was selected to test the relationship hypotheses. The correlation estimates are listed in table 1, with significant coefficients flagged with *. These results are critically analyzed below.

The first desirable observation from table 1 was very few significant correlations between the independent PM factors (age, gender, experience, education and certification), with exception to age and experience. This was a desired result because it showed the independent factors were not likely going to interfere with one another as mediators of the dependent variables. The *a priori* literature suggested age would often be positively correlated with experience because older PMs have more experience, which was the finding in the current study (*rho* = 0.758, *p* < .001). There was also a significant small bivariate correlation between gender and education (*rho* = -0.44, *p* < .05) which we could interpret as females in the sample had a slightly higher education level.

Another desirable result in table 1 was age was not related to risk management decision making in the basic condition (*rho* = 0.323, *p* > .05) but the opposite occurred for the bias treatment (*rho* = 0.538, *p* < .01). Due to this we cannot fully accept the hypothesis H1a: PM age is related to project result because age was related to the biased decision. Perhaps older PMs had significant insight which came to play when making a biased risk management decision in a crisis as contrasted with a basic no incentive business-as-usual situation.

Gender was not related to risk management decision making (*rho* = -0.153, *p* > .05 and *rho* = -0.246, *p* > .05). This finding does not support hypothesis H1b: PM gender is related to project result, so it was rejected. However, certification was related to risk management decision making (*rho* = 0.549, *p* < .01 and *rho* = 0.712, *p* < .001). This finding supports hypothesis H1c: PM certification is related to project result so it was accepted. This indicates PMs who were certified generally made better risk management decisions in the basic and biased conditions.

Education was not related to project result (*rho* = 0.052, *p* > .05, *rho* = 0.017, *p* > .05) for either risk management decision. Therefore, H1d: PM education is related to project result was rejected. Likewise, experience was not related to project result (*rho* = 0.226, *p* > .05, *rho* = 0.376, *p* > .05). We therefore rejected hypothesis H1e: PM experience is related to project result.

Competency was significantly related to both risk management decisions (*rho* = 0.680, *p* < .001 and *rho* = 0.681, *p* < .001), which was a desirable finding. Based on that result we accepted hypothesis H1f: PM competency is related to project result. We also noticed a non-hypothesized significant correlation between certification and competency (*rho* = 0.688, *p* < 0.001). This suggests PMs who were certified were also more competent at least based on the single question test we gave them during the experiment. Caution should be used to assume this single test was a comprehensive indicator of competency as it was a cursory validation for experimental purposes.

There were interesting correlation results between the two dependent variables for the risk management decision making scenarios. This will not answer the second set of hypotheses but it can illustrate if the respondent's behavior in the basic and bias conditions were similar. The correlation was significant between the basic and biased risk management decision making (*rho* = 0.610, *p* < .001) which portrays a moderate consistency in the respondent's behavior across both treatments. We can summarize the first set of hypotheses as being only certification and competency had significant impacts on risk management decision making. Therefore, only those two factors will be tested further.

4.3 Second Set of Hypothesis Test Results

In order to answer the second set of hypotheses, a multivariate comparative test was needed was each hypothesis referred testing the risk management decision making behavior of subgroups of respondents and there were two dependent variables representing the experimental conditions ─ the basic and biased project decision result in a crisis. Since there were two independent factors representing the certification and competency subgroups as well as two dependent variables (the two conditions basic and biased), this required a MANOVA. The objective was to test if each subgroup (certified or not certified as well as competent or not competent) correctly cancelled the failing project, in order to answer the RQ and second set of hypotheses. As noted earlier, the experiment was configured to represent a failing project which according to the displayed standard operating procedures and organizational policy, it must be cancelled. The biased treatments represented artificial lures to persuade the PM to make an incorrect decision if more money was offered and the prospect of continuing the project seemed possible.

First a Shapiro-Wilk test for multivariate normality was conducted, which was to ensure the assumption of normality was met for MANVOVA, otherwise a nonparametric Friedman's MANOVA must be utilized. This was different than the univariate SW test performed earlier because all of the factors and both of the dependent variables were included this time. The result was favorable with *SW* = 0.927, (p = 0.082). Next the MANOVA was run with the key estimates presented in table 2. The DF column refers to degrees of freedom, the *F* is the approximate *F-test* estimate, the *Pillai* is the multivariate *Pillai* trace significance test result (which is somewhat similar to a *t-test* but in a multivariate context). The numerator (num.) *DF*, and denominator (den.) *DF*, for the Piallai trace test is listed, followed by the *p*-value.

The MANOVA results in table 2 were desirable because there were significant differences between both subgroups of respondents, for certified and competent, in terms of their making the correct risk management decision in both experimental conditions. Certified versus not certified respondents were significantly different in their risk management decision across both dependent variables (*F* (Bohnert et al., 2019; Pace, 2019) = 17.824, *Pillai* (Ghossein et al., 2018; Prewett & Terry, 2018) = 0.652, $p <$ .001). The risk management decisions for the failing project of respondents who passed the competency test were significantly different that those who did not (*F* (Bohnert et al., 2019; Pace, 2019) = 3.723, *Pillai* (Ghossein et al., 2018; Prewett & Terry, 2018) = 0.282, p = .043). However, there was significant interaction between these 2 x 2 factorial experimental conditions, as observed by the certified * competent terms tested in the MANOVA (*F* (Bohnert et al., 2019; Pace, 2019)= 11.465, *Pillai* (Ghossein et al., 2018; Prewett & Terry, 2018) = 0.547, $p <$ .001). To further check this, a Box's M-test for homogeneity of the covariance matrices was executed, with a significant result (χ^2 (Carlton & Peszynski, 2018) = 39.74, $p <$ .001). This indicates there were some differences in the covariance correlation matrices, and considering the earlier significant SW test for the biased decision, the second dependent variable likely did not approximate a normal distribution. This was not a series impediment to the current study because this was a human behavior experiment with high complexity in a simulated crisis and considerable monetary persuasion in the last treatment so deviations in the second biased decision dependent variable was expected. However, additional tests were conducted to answer the second set of hypotheses.

TABLE 1: DESCRIPTIVE SAMPLE ESTIMATES AND SPEARMAN CORRELATION COEFFICIENTS

Variable	Mean	Age	Gender	Experience	Education	Certified	Competent	Basic
Age	42.458							
Gender	0.708	-0.186						
Experience	10.958	0.758 ***	-0.266					
Education	3.375	0.288	-0.440 *	0.386				
Certified	0.5	0.163	-0.092	0.085	-0.030			
Competent	5.458	0.390	-0.100	0.249	0.166	0.688 ***		
Basic Decision	3.688	0.323	-0.153	0.226	0.052	0.549 **	0.680 ***	
Biased Decision	2.891	0.538 **	-0.246	0.376	0.017	0.712 ***	0.681 ***	0.610 **
* p < .05, ** p < .01, *** p < .001								

TABLE 2: MANOVA EXPERIMENTAL SUBGROUP COMPARISON OVERALL FACTOR ESTIMATES

Hypothesis groups	*DF*	*F*	*Pillai*	Num. *DF*	Den. *DF*	*P*
(Intercept)	1	186.309	0.951	2	19	<.001
Certified	1	17.824	0.652	2	19	<.001
Competent	1	3.723	0.282	2	19	0.043
Certified * Competent	1	11.465	0.547	2	19	<.001
Residuals	20					

TABLE 3: REPEATED MEASURES ANOVA EXPERIMENTAL SUBGROUP DETAILED FACTOR ESTIMATES

Within subjects effects (overall)	SS	DF	Mean SS	F	P	η²
Good decision	13.377	1	13.377	50.340	< .001 *	0.160
Residuals	5.315	20	0.266			
Between subjects effects						
Competent	23.257	1	23.257	17.091	< .001*	0.278
Certified	5.629	1	5.629	4.137	0.055 °	0.067
Certified * Competent	0.455	1	0.455	0.334	0.570	0.005
Residuals	27.214	20	1.361			
* p < .05, ** p < .01, *** p < .001; SS = Type III SS; ° accepted as p < .05 in an exploratory context.						

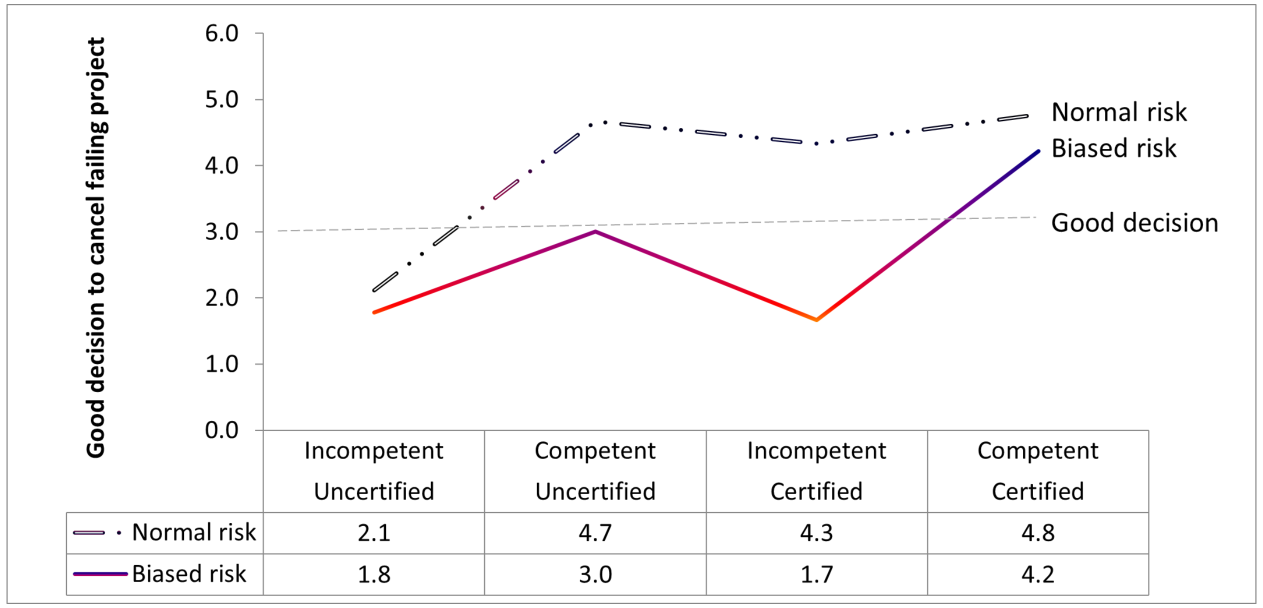

FIGURE 1: PM RISK MANAGEMENT DECISION MAKING ANALYSIS BY COMPETENCY AND CERTIFICATION SUBGROUPS

The parametric ANOVA repeated measures technique (or nonparametric Friedman's test when normalcy tests fail) was appropriate to answer the RQ and second set of hypotheses by determining which subgroups (competent and or certified) made the correct decision to cancel the failing project in the experiment. On a preliminary basis we confirmed the normalcy of the two dependent variables (the PM's decision to cancel a failing project on the 1-5 ratio scale with 5 being the best choice. The Levene test for equality of variances was significant for both dependent variables (*F* (Bohnert et al., 2019; Kurek et al., 2017) = *4.143,* p = 0.019, *F* (Bohnert et al., 2019; Kurek et al., 2017) = 6.387, p = .003). This indicates the assumption of within group variance was violated, so as a precaution Friedman's test was run in parallel with ANOVA.

As noted above, since this was an experiment with considerable monetary persuasion treatments in a complex risk decision making, significant variance was expected within and between subgroups. The result of the Friedman test was as desired for the risk management decision to cancel the failing project (x^2 (Pace, 2019) = 6.000, Kendall's *W* = 0.805, p = .014). This indicated there were significant differences in making the correct risk management decision to cancel the failing project between the PM subgroups of competency and certification.

The repeated measures ANOVA were executed to calculate the detailed subgroup partial coefficients and effects, with standard error captured across all factors and variables. The purpose was to answer the RQ and the second set of hypotheses. The results were very desirable in terms of supporting the second set of hypotheses. The key estimates are summarized in table 3. In table 3, the

SS refers to sum of squares, and η^2 is an approximate factor partial effect size in a multivariate model. The other headings were explained earlier for table 2.

Overall, the results of the experiment were as hypothesized being that there were statistically significant differences in risk management decision making for a failing project and PM's with higher competence made better decisions, when all other known factors were controlled or accounted for. As shown in table 3 at the top, the risk management decision to correctly cancel the failing project (good decision) was significant (*F* (Bohnert et al., 2019; Pace, 2019) = 50.34, p < .001, η^2 = 0.16) having an approximate 16% effect size. This means there were significant differences but a post-hoc test is required to confirm this, then the means must be examined to answer the second set of hypotheses to determine which subgroup performed better.

Another desirable result from table 3 was the overall interaction terms were insignificant. There was a small overall shared variation on the factors and dependent variables at approximately 6% which we accepted due to the exploratory nature of the experiment — this slight interaction was likely due to the repeated treatments, as respondents would have remembered the first treatment when making the second risk management condition. The between subjects' section of table 3 explains the comparisons. A certified respondent made a significantly different decision as compared to a PM who was not (certified F [1,20] = 17.091, p < .001, η^2 = 0.278) with an approximate 28% factor effect size. Similarly, a competent respondent made a significantly different decision as compared to one who failed the competency test (certified *F* (Bohnert et al., 2019; Pace, 2019) = 4.137, p < .001, η^2 = 0.278) with an approximate 7% factor effect size. Again, there were no significant interactions of the two independent factors on the dependent variables, which was more important to consider as compared to the earlier small shared variation.

An ANOVA post hoc subgroup analysis using pairwise t-tests confirmed the competent and certified groups made significantly different risk management decisions in the simulated project crisis (mean difference = 1.219, *SE* =0.172, t = 7.095, Cohen's D = 1.448, p < .001, where SE refers to standard error). In order to answer the second set of hypotheses, the acceptable mean must be acknowledged, which would be 3 or higher on the 1-5 ratio scale. Any risk management decision at or above a 3 would be considered successfully cancelling a failing project. The experiment was purposefully setup to present scenarios of a failing project in a crisis so the only acceptable decision according to the organization's policy displayed to the respondents was to cancel the project regardless of financial incentives. This was made clear to respondents.

In the ANOVA post hoc mean test results, PM's who were considered competent scored at least a 3 and significantly higher than those who were not competent in the basic risk management decision (competent M = 4.71 vs. not competent M = 2.63). Subsequently, hypothesis H2b was accepted, that competent PM's are more likely to cancel a failing project without bias (basic decision) was accepted. Similarly, PMs who were competent scored higher than 3 and significantly higher than those who were not competent in the biased risk management decision (competent M = 3.91 versus not competent M = 1.81). Consequentially, hypothesis H2d was accepted, that competent PM's are more likely to cancel a failing project with bias (biased decision).

In the ANOVA post hoc means analysis, PMs who were certified scored significantly higher than those not certified and above 3 referring to a good risk management decision to cancel a failing project (certified M = 4.58 versus uncertified M = 2.76). Based on this result there was support to accept the hypothesis H2a that certified PM's are more likely to cancel a failing project without bias (basic decision). In the second treatment condition where incentives were added, certified PMs answered above 3 and there were statistically higher than those uncertified (certified M = 3.55 versus uncertified M = 2.17). Accordingly, hypothesis H2c was accepted, being that certified PM's are more likely to cancel a failing project with bias (biased decision).

In summary, all hypotheses were supported and the RQ was answered. The MANOVA and repeated measures ANOVA on the experimental data established that both certified and competent PM's were likely to make better risk management decisions to cancel a failing project in a crisis without bias and with bias conditions.

5. CONSLUSIONS AND OUTLOOK

Turning back to our rationale for initiating this study, the objective was to use a rigorous controlled experiment simulating a project failure to measure how individual PM factors impacted risk management decision making in a crisis while controlling for other contextual elements including project and firm level characteristics. We successfully answered the RQ of how much does bias impact PM risk management decision making in a crisis: Bias significantly impacts the PM decision, but certification and competency resulted in the best decisions.

In summary, PM competency and certification were significantly and positively related to making a good risk management decision during a crisis in both the basic as well as the biased circumstances. In fact, PM competency trumped certification (and all other measured factors) in this regard. The experimental data indicated that PM's with higher risk management competency levels made better decisions in both basic as well as conditions biased with high prospect value (extra money and ego protection).

We can further examine the experimental data to prove this. Figure 1 visualizes the conditional means (M) of the basic versus biased risk management decision making to cancel a failing project by the PM certification level and competency test pass/fail score. The data in figure 1 was calculated on subgroups using joint conditional probability logic — for example each cell contains the mean for a PM with competency = 1 and certified = 1, for the basic risk management decision, and another row for the biased condition, and so on for all other combinations. A perfect decision on the y-axis would be a 4-5, but as noted earlier, a 3 would be barely acceptable indicating unsure (giving the respondent the benefit of the doubt they would cancel if given more time to contemplate the ethical impacts), while the 1-2 scale was clearly a poor decision meaning to not cancel.

In figure 1, the biased experimental condition was identical to the basic risk decision context except that a significant value prospect incentive was offered to not cancel the project. In other words, economic and emotional bribes were proposed to see if the PM would overlook the required risk decision to cancel the failing project, which was clearly explained as the right course of action based on the standard operating procedures.

Figure 1 clearly illustrates the theoretical problem with PMs, at least in the experimental sample. The x-axis and integrated data value table indicates the repeated measures subgroup combinations while the y-axis represents the same risk management decision making project result scale from 1-5 where 3 was the minimum acceptable and 5 was highest. The left to right straight dashed line represents the minimum acceptable scale value of 3 for illustrative purposes. The data series dotted line plots the normal risk management treatment responses while the solid line shows the biased condition results.

Figure 1

The interpretation of figure 1 is that incompetent and uncertified PM's made the worst decisions by continuing the failing project regardless of risks or the organization policy. Competent but uncertified PM's (the second subgroup combination in figure 1) made the acceptable risk management decisions during the crisis, especially in the normal treatment (M = 4.7) and also just meeting the minimum benchmark in the biased condition (M = 3.0). Here is where the hypothetical problem occurred. A certified but incompetent PM made a good risk management decision in the basic scenario (M = 4.3) but unfortunately the PM failed to cancel the failing project in the biased condition (M=1.7). A competent and certified PM made the nest risk management decisions in both the basic and biased conditions (M = 4.8, M = 4.2).

We would expect a professional certified PM to cancel a failing project when a severe risk event occurred and organizational procedures in the experiment mandated that. What was not expected was a certified PM could be unduly influenced by money and potential success as to go against training and overlook organizational policies. To generalize this further is money and ego that powerful they could influence even a certified PM to perform unethically? This was certainly not the result for a competent but uncertified PM who made the correct decisions in both the basic and biased conditions. This could potentially become a significant problem for recruiters if certification ranks high in the hiring criteria but no evidence of PM competency is sought. On the other hand, this problem may be solvable. Certified PM's could receive additional training to become more competent for risk management. Actually, risk management is one of the newer knowledge areas in the project management discipline so perhaps even certification courses could be revised to include experiential learning exercises for quantifying internal risks (e.g., at the resource and milestone level) as well as mitigating external risk events.

Competence seemed to overrule certification in the experiment. Competency had higher significant correlations to making good risk event decisions. Competent but uncertified PM's made the better decisions, but still not perfect though. The PM did cancel the failing project in the basic scenario (M=4.7) but they were undecided about cancelling the failing project in the biased value prospect condition (*M*=3). If we extend the benefit of the doubt here, perhaps the competent but uncertified PM's would elect to make the correct decision to cancel the failing project in the biased condition if they were given longer to think about it (only 15 minutes was allowed). Conversely, we have no proof the decision could go the other direction with more time! This begs the question: Are money and ego strong enough to overpower ethical risk management decision making in competent and certified PM's? Ethics and code of conduct are core knowledge areas of professional PM certification so it seems possible the lack of certification could be the root cause. If so, employers could address this problem by endorsing PM certification.

Competency and certification clearly produced the best decision. Competent and certified PM's made the best decisions in all experimental conditions, with the PM cancelled the failing project in both the basic (*M* = 4.8) and biased conditions (*M* = 4.2). Clearly certification seems to result in better decisions when there was no bias, but this does not help a PM who was incompetent to make the correct decision. Something is clearly at play here. It appears bias, especially economic and emotional gains, as designed into the experimental condition, are strong influencers working against the ethical judgment of PM's who are certified but could not score well in the competency test. This may also be a solvable problem. If ethics is the underlying problem, despite the PM being certified, then obviously more ethics and or organizational training would be needed. Perhaps employers could offer short online brush up courses similar to what many of us PMs regularly take to comply with OSHA regulations.

In summary, we argue these findings ought to generalize globally as a plausible factor causing some of the projects to fail, if tested further with larger experiments. More research will be needed to further explore these very interesting experimental findings within other countries and for different cultures.

5.1 Limitations and Future Research Outlook

First and foremost, the key limitation is that generalizations must be cautioned because this was a controlled experiment with 24 participants based in the U.S. Although experienced PM's were randomly selected, we have to keep in mind that certification was self-reported and competency was judged solely by the score on one open book test question during the experiment to calculate risk using the best-practice PERT risk management methodology (which an experienced and competent PM ought to know).

Additionally, all organization level factors were controlled, on purpose, as they were defined in the experimental context according to what the authors wanted. This means this experimental project context would not necessarily be equivalent to that faced in the real world nor would the organizational conditions be identical. Controlling the factors made the experiment very successful and manageable but at the cost of reducing generalizability.

The bias conditions in our experiment were a combination of money and implied ego – there are other bias constructs which could be tested in future experiments. We suggest other researchers do that, test additional types of bias on risk management decision making.

Finally, we can point out that the yes/no conditions for PM competency were calculated from the actual test scores to facilitate applying repeated measures ANOVA which was in turn driven by our RQ and hypotheses. As explained earlier we recorded the competency score from the test as a 0-10 continuous ratio data type but we transformed this to a binary field (60% being a pass or 1 and 0 for lower scores). This was done to also facilitate using MANOVA when competency was positioned as a subgroup factor. This was similar to what one would encounter when designing 2x2 factorials or when using the general linear model or mixed ANOVA statistical techniques because continuous data types cannot be easily processed as fixed factor indicators for group comparisons. Therefore, we recommend additional experiments using MANCOVA, a multivariate technique able to handle continuous covariates with ordinal or nominal factors and multiple dependent variables.

We admit our results were experimental because all project, organization and global factors were controlled - they were static — so those were not tested. Nevertheless, the current study findings revealed significant statistical evidence to substantiate our hypotheses that cognitive bias altered PM's risk management decision making more than other individual factors such as experience, certification or competence.

A repeated measures controlled experiment is the gold standard of scientific high power techniques for human behavior testing. In the current study, the repeated measures-controlled experiment was rigorous because it took place in approximately adjacent points in time and each participant received all treatments. Most importantly, each PM received the same identical competency test and both risk management decision making experimental treatments. The reason this repeated measures-controlled experiment was scientifically robust was that all variables not observed or captured were controlled by making them static and declared to participants. In this way, the PM individual differences remained the only varying factors when everything else was controlled, thereby eliminating isolated subgroup interaction between specific conditions and individuals.

However, the current study was limited to a quasi-experiment since it was conducted asynchronously online (not in a lab) and while the PMs were randomly sampled they self-selected to participant. To lessen that affect, all aspects of the experiment were confidential to participants except for the declared static variables, the duration was one week, and participants had no contact with one another. It was also possible that lingering sentiments after the first treatment may have cognitively impacted the second treatment, since it was a repeated measures design by purpose.

Thus, in closing, we recommend caution for generalizing the results and we encourage replication as well as extension of this. To that end we have made our experimental data available for other researchers by contacting the PI first author or by download the comma delimited file from http://kennethstrang.com/data/ PM_decision_making_experiment24-strang(2022).csv. We encourage researchers to improve upon these findings and share their results.

REFERENCES

Anthopoulos, L., Reddick, C. G., Giannakidou, I., & Mavridis, N. (2016). Why e-government projects fail? An analysis of the Healthcare. gov website. *Government information quarterly, 33*(1), 161-173. https://doi.org/10.1016/j.giq.2015.07.003

Barrows, M., Clevenger, C. M., Abdallah, M., & Wu, W. (2020). Value of certifications when seeking construction employment. *International Journal of Construction Education and Research, 16*(1), 61-79. https://doi.org/10.1080/15578771.2019.1575936

Bohnert, A., Gatzert, N., Hoyt, R. E., & Lechner, P. (2019). The drivers and value of enterprise risk management: evidence from ERM ratings. *The European Journal of Finance, 25*(3), 234-255. https://doi.org/10.1080/1351847X.2018.1514314

Borbath, M., Blessner, P., & Olson, B. (2019). An empirical approach to the evaluation of defence contractor performance. *International Journal of Logistics Research and Applications, 22*(2), 138-153. https://doi.org/10.1080/13675567.2018.1497148

Brandon, T. P., Allen, B. M., & Menefee, S. J. (2022). Which Experience Factors Predict Overall Expertise and Percent Success? *The Journal of Modern Project Management, 9*(3), 20-31. https://journalmodernpm.com/index.php/jmpm/article/download/JMPM02802/8

Callahan, C., & Soileau, J. (2017). Does enterprise risk management enhance operating performance? *Advances in accounting, 37*, 122-139. https://doi.org/10.1016/j.adiac.2017.01.001

Carlton, D., & Peszynski, K. (2018). Situational Incompetence: the failure of governance in the management of large scale IT projects. *International Working Conference on Transfer and Diffusion of IT* (pp. 224-244). Springer. https://doi.org/10.1007/978-3-030-04315-5_16

Conner, M., Godin, G., Sheeran, P., & Germain, M. (2013). Some feelings are more important: cognitive attitudes, affective attitudes, anticipated affect, and blood donation. *Health Psychology, 32*(3), 264–272. https://doi.org/10.1037/a0028500

CSOTC. (2017). *Enterprise Risk Management: Aligning Risk with Strategy and Performance*. New York, NY: Committee of Sponsoring Organizations of the Treadway Commission (CSOTC).

Damoah, I. S., Akwei, C. A., Amoako, I. O., & Botchie, D. (2018). Corruption as a source of government project failure in developing countries: Evidence from Ghana. *Project Management Journal, 49*(3), 17-33. https://doi.org/10.1177%2F8756972818770587

Eckerd, A., & Snider, K. (2017). Does the program manager matter? New public management and defense acquisition. *The American Review of Public Administration, 47*(1), 36-57. https://doi.org/10.1177%2F0275074015596376

Elbanna, S., & Child, J. (2007). The influence of decision, environmental and firm characteristics on the rationality of strategic decision-making. *Journal of Management Studies, 44*(4), 561-591. https://doi.org/10.1111/j.1467-6486.2006.00670.x

Ghossein, T., Islam, A. M., & Saliola, F. (2018). *Public Procurement and the Private Business Sector: Evidence from Firm-Level Data*. World Bank. https://doi.org/10.1596/1813-9450-8575

Goodwin, Y., & Strang, K. D. (2012). Socio-cultural and multi-disciplinary perceptions of risk. *International Journal of Risk and Contingency Management (IJRCM), 1*(1), 1-11. http://doi.org/10.4018/ijrcm.2012010101

Goyal, P., Gupta, P., & Yadav, V. (2021). Antecedents to heuristics: decoding the role of herding and prospect theory for Indian millennial investors. *Review of Behavioral Finance*. https://doi.org/10.1108/RBF-04-2021-0073

Hughes, D. L., Rana, N. P., & Simintiras, A. C. (2017). The changing landscape of IS project failure: an examination of the key factors. *Journal of Enterprise Information Management, 30*(1), 142-165. https://doi.org/10.1108/JEIM-01-2016-0029

Jennings, W., Lodge, M., & Ryan, M. (2018). Comparing blunders in government. *European Journal of Political Research, 57*(1), 238-258. https://doi.org/10.1111/1475-6765.12230

Jordão, A. R., Costa, R., Dias, Á. L., Pereira, L., & Santos, J. P. (2020). Bounded rationality in decision making: an analysis of the decision-making biases. *Verslas: teorija ir praktika, 21*(2), 654-665. http://dx.doi.org/10.3846/btp.2020.11154

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-292. https://doi.org/10.2307/1914185

Klein, D. E., Woods, D. D., Klein, G., & Perry, S. J. (2016). Can we trust best practices? Six cognitive challenges of evidence-based approaches. *Journal of Cognitive Engineering and Decision Making, 10*(3), 244-254. https://doi.org/10.1177%2F1555343416637520

Korstanje, M. E., Strang, D. K., Tzanelli, R., & Moreno Garcia, R. (2020). Turismo, riesgo y cambio climático: Un camino alternativo. *Estudios y perspectivas en turismo, 29*(1), 214-227. http://www.estudiosenturismo.com.ar/PDF/V29/N01/v29n1a12%20res.pdf

Kurek, E., Johnson, J., & Mulder, H. (2017). Measuring the value of Enterprise Architecture on IT projects with CHAOS Research. *Systems, Cybernetics, and Informatics, 15*, 13-18. http://dx.doi.org/10.13140/RG.2.2.21487.12966

Masticola, S. P. (2007). A simple estimate of the cost of software project failures and the breakeven effectiveness of project risk management. *2007 First International Workshop on the Economics of Software and Computation* (pp. 6-6). IEEE. https://doi.org/10.1109/ESC.2007.1

McShane, M. (2018). Enterprise risk management: history and a design science proposal. *The Journal of Risk Finance, 19*(2), 137-153. https://doi.org/10.1108/JRF-03-2017-0048

O'Boyle, E., Banks, G. C., Carter, K., Walter, S., & Yuan, Z. (2019). A 20-year review of outcome reporting bias in moderated multiple regression. *Journal of Business and Psychology, 34*(1), 19-37. https://doi.org/10.1007/s10869-018-9539-8

Pace, M. (2019). A correlational study on project management methodology and project success. *Journal of Engineering, Project, and Production Management, 9*(2), 56-65. http://dx.doi.org/10.2478%2fjeppm-2019-0007

Prewett, K., & Terry, A. (2018). COSO's updated enterprise risk management framework—A quest for depth and clarity. *Journal of Corporate Accounting & Finance, 29*(3), 16-23. https://doi.org/10.1002/jcaf.22346

Raftery, J. (1994). *Risk Analysis in Project Management*. First UK: E & FN Spon An imprint of Routledge.

Tapas, P., & Pillai, D. (2021). Prospect theory: an analysis of corporate actions and priorities in pandemic crisis. *International Journal of Innovation Science*. https://doi.org/10.1108/IJIS-10-2020-0241

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty, 5*(4), 297-323. https://doi.org/10.1007/BF00122574

ABOUT AUTHORS

Professor Kenneth David Strang, Doctorate, MBA, BS, AS, FLMI, CNA, CSCS, CRP, CPP

Professor Strang has almost 300 publications (http://kennethstrang.com) and at the time of writing he has consistently ranked in the 95-97 percentiles of world-wide scholars at Research Gate. Dr. Strang has been working since 1983. Ken has been employed by W3-Research, a business analytics company, since 2010. He is a retired full professor who occasionally lectures at reputable universities and associations. Ken has a professional Doctorate in project management operations research, an MBA in strategic management, a BS in marketing, and an AS in computer science all with summa cum lade high grade distinctions. He has six professional certification/licenses: Certified Netware Administer (security), a Fellow Life Management Institute (insurance actuary), a Certified Research Professional (scientific research), a Project Management Professional, a Certified Supply Chain Specialist/Procurement Professional, and a World Medical Association research ethics certification. Ken conducts research in business, healthcare, logistics/supply chain, manufacturing, marketing, mining, new product development, leadership, organizational behavior, industrial psychology, tourism marketing, human resource management, learning styles, online learning, nuclear-hydro-solar-wind clean energy production, and procurement project management. He edited a best-selling research design-methods textbook for Palgrave-Springer and he won over $8M USD in grant-proposals to support his clients. He also won two Emerald Literati Awards. He has managed over 115 studies, taught over 1000 multicultural students with an anonymous course opinion survey satisfaction median of 5/5 and a mean of 4.5/5. Ken volunteers for many charities and not-for-profit social advocate associations.

Associate Professor Narasimha Rao Vajjhala, PhD, MSCS, MBA, PMP

Dr. Narasimha Rao Vajjhala is working as an Associate Professor and the Chair of the Information Systems department at the School of IT and Computing at the American University of Nigeria. He had previously worked at the Faculty of Engineering and Architecture at the University of New York Tirana, Albania. He is a senior member of ACM and IEEE. He is the Editor-in-Chief for the *International Journal of Risk and Contingency Management*, and a former Managing Editor. He is a member of the Risk Management Society, and the Project Management Institute. He has over 20 years of experience in teaching programming, design, security, analytics and database-related courses at both graduate and undergraduate levels in Europe and Africa. He has also worked as a consultant in technology firms in Europe and has experience participating in EU-funded projects. He has completed a Doctorate in Information Systems and Technology (United States); he holds a Master of Science in Computer Science and Applications (India), and a Master of Business Administration with a specialization in Information Systems (Switzerland).