Expert Judgment in Project Management

Narrowing the Theory-Practice Gap

Paul S. Szwed, DSc, PMP

Project Management Institute

Expert Judgment in Project Management: Narrowing the Theory-Practice Gap

Paul S. Szwed, DSc, PMP
Professor
Department of International Maritime Business
Massachusetts Maritime Academy

Library of Congress Cataloging-in-Publication Data has been applied for.

ISBN: 978-1-62825-116-6

Published by: Project Management Institute, Inc.
14 Campus Boulevard
Newtown Square, Pennsylvania 19073-3299 USA
Phone: +610-356-4600
Fax: +610-356-4647
Email: customercare@pmi.org
Internet: www.PMI.org

PMI Publications welcomes corrections and comments on its books. Please feel free to send comments on typographical, formatting, or other errors. Simply make a copy of the relevant page of the book, mark the error, and send it to: Book Editor, PMI Publications, 14 Campus Boulevard, Newtown Square, PA 19073-3299 USA.

To inquire about discounts for resale or educational purposes, please contact the PMI Book Service Center.

PMI Book Service Center
P.O. Box 932683, Atlanta, GA 31193-2683 USA
Phone: 1-866-276-4764 (within the U.S. or Canada) or +1-770-280-4129 (globally)
Fax: +1-770-280-4113
Email: info@bookorders.pmi.org

The paper used in this book complies with the Permanent Paper Standard issued by the National Information Standards Organization (Z39.48—1984).

10 9 8 7 6 5 4 3 2 1

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to acknowledge the continuous and ongoing support of my wife, Anita. Thank you!

This study represents the realization of a research goal developed at an early stage in my dissertation work almost 15 years ago. Even before embarking on my doctoral work as a risk analyst, operations manager, and project management practitioner, I observed firsthand that the practice of expert judgment elicitation was frequently conducted in an ad hoc manner. I knew this practice was fraught with potential errors and biases that would lead to less than desired estimations, forecasts, and predictions. Over the years, I have been able to develop guidelines for the organizations I have been associated with and have also contributed to the development of international standards in the area. However, through this research and by the publication of this monograph, I am hopeful that an entire profession will begin to become more adept at eliciting expert judgment—to start to close the theory-practice gap.

This study is funded primarily through a research grant from the Project Management Institute (PMI), whose support is greatly appreciated. I am also thankful for the guidance of PMI's Manager of Academic Resources, Carla Messikomer, and V. K. Narayan (my assigned research mentor for this project), as well as the entire staff at PMI, who supported my efforts (especially Kristin Dunn and Kimberly Whitby).

I am also grateful for the in-kind funding and support I have received from the Massachusetts Maritime Academy. I would also like to thank Mohammed Marzuq and Liz Novak for their assistance on phase 1, and Mason Fortier and Kate McLaren for their assistance on phase 2.

Finally, the views expressed in this report are those of the author and do not necessarily represent those of the Project Management Institute or the Massachusetts Maritime Academy.

Executive Summary

Problem: In the face of unknown futures, project managers often turn to expert judgment as a source of key information in an effort to ensure that projects are completed on time, on budget, and in accordance with stakeholder expectations. As a project management tool/technique, expert judgment is ubiquitous in the body of knowledge. However, most organizations do not have written guidance on how to elicit expert judgment, and most project managers rely on ad hoc methods for gathering expert judgment, which is known to result in flawed information that can adversely impact project success. Therefore, this study sets out to identify ways in which expert judgment might be improved for project management.

Background: This study was primarily sponsored by the Project Management Institute. The study comprised three phases. The first phase investigated the state of the art/science of expert judgment elicitation broadly (e.g., across a variety of disciplinary areas, such as engineering, environmental management, medicine, political science, and space exploration). The second phase identified the state of the practice of expert judgment in project management. Together, the first two phases identified several theory-practice gaps. The third phase examined methods for closing key gaps. A complete description of the study (including the problem, the background, and the methods) is provided in Chapter 1 of this report.

Process: Spanning 15 months, this study employed a mixed-methods approach. A comprehensive review of the literature was conducted to identify the state of the art/science. Using a 10-step process, more than a thousand relevant articles and studies were found for the review period, of which dozens were deemed applicable to the problem at hand. Chapter 2 of the report contains a detailed description of phase 1. A descriptive survey was conducted to determine the state of the practice. Surveys were sent to all of the regional chapters of the Project Management Institute, a professional organization with nearly half a million members. There were more than 400 responses from a representative sample of the organization. The sample was subjected to descriptive statistics and multivariate analysis. Chapter 3 of the report provides a detailed description of phase 2. Finally, to close a key gap, experiments were conducted. Undergraduate students from two universities provided expert judgments through tested elicitation protocols. The details of phase 3 are described in Chapter 4 of the report.

Findings and Conclusions: There were many findings and conclusions. In general, the state of the art/science of expert judgment outside of project management is established and continually evolving. The state of the practice of expert judgment in project management is informal and emergent. There is considerable opportunity to inform and mature the practice of expert judgment in project management by adopting best practices from other professions and disciplines. Further specific findings and the rationale for them are provided throughout the report and in Chapter 5 specifically.

Recommendations: To advance the practice of using expert judgment in project management, the following general recommendations are provided:

- Organizations should provide written guidance on how to conduct expert judgment using a standard framework that includes the following seven steps: frame the problem, plan the elicitation, select the experts, train the experts, elicit the judgments, analyze and combine the judgments, and document and communicate the results.
- In order to best leverage the judgment of experts, clearly frame the problem to be considered and identify the exact nature of the information to be sought.

- Determine if the expert judgment is generative (i.e., creating lists, risks, options, etc.) or evaluative in nature (i.e., estimating cost, duration, quantifying a phenomenon of interest, etc.). Select an appropriate expert judgment elicitation method. Create an elicitation protocol using that method and test it with normative experts.
- Using appropriate criteria, select a diverse pool of four to eight experts from both inside the project management organization and outside it, including key stakeholders. Fluency (numeracy) tests will help in selecting experts for generative (evaluative) tasks.
- Train experts prior to the elicitation on the reasons judgments are needed, the process of the protocol, and means to mitigate known biases through practice.
- Elicit expert judgments using previously selected, established methods. Do not over-rely on brainstorming or ad hoc processes.
- Where possible, evaluate expert performance in order to weight or select judgments accordingly. Unless there is a compelling reason to do otherwise, use simple averaging to combine evaluative judgments. Use an interactive consensus process to combine generative judgments.
- Document the entire elicitation process, including recording the method used, expert information, and resulting judgments. This will serve as an organizational process asset and lessons learned for future projects that involve the elicitation of expert judgment.

Further specific recommendations are provided in Chapter 5 of the report, and the References list is a good source for self-study to improve proficiency in conducting expert judgment elicitations in project management.

CHAPTER 1

Introduction

Given the temporary and unique nature of projects (typically within new and different contexts and conditions), the practice of project management is often a complex and challenging endeavor. Project managers must be able to deal with uncertainty and unknowns. Without access to historic data or known information, project managers use estimation, forecasting, and prediction to plan for projects. One of the most commonly used approaches is to gather expert judgment to "fill in the gaps" in information. For example, project managers may obtain the judgment of experts to estimate the resources necessary to successfully complete a project. Likewise, project managers may gather expert judgment to forecast likely future scenarios or risks that may occur during the life of a project.

1.1 Background

Expert judgment is by far the most frequently listed tool/technique in *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)* – Fifth Edition. As illustrated in Table 1, expert judgment is explicitly listed as a tool/technique for 28 of the 47 project management processes (59.6%) and mentioned implicitly in another six processes (bringing the frequency to 72.3%).

For example, within the *PMBOK® Guide*, expert judgment is suggested as a potential tool/technique in *all* six of the processes contained in the Project Integration Management Knowledge Area. On the other end of the spectrum, expert judgment is not

Table 1 *PMBOK® Guide* processes that list expert judgment as a tool/technique.

Knowledge Areas	*Group Processes* 1	2	3	4	5	6	7
4. Project Integration Management	Directly Listed	Directly Listed	Directly Listed	Directly Listed	Directly Listed	Directly Listed	
5. Project Scope Management	Directly Listed	Indirectly Listed	Directly Listed	Directly Listed	Indirectly Listed	Not Listed	
6. Project Time Management	Directly Listed	Directly Listed	Not Listed	Directly Listed	Directly Listed	Not Listed	Indirectly Listed
7. Project Cost Management	Directly Listed	Directly Listed	Directly Listed	Indirectly Listed			
8. Project Quality Management	Not Listed	Indirectly Listed	Not Listed				
9. Project Human Resource Management	Directly Listed	Indirectly Listed	Not Listed	Not Listed			
10. Project Communications Management	Not Listed	Not Listed	Directly Listed				
11. Project Risk Management	Directly Listed	Directly Listed	Directly Listed	Directly Listed	Directly Listed	Not Listed	
12. Project Procurement Management	Directly Listed	Directly Listed	Not Listed	Not Listed			
13. Project Stakeholder Management	Directly Listed	Directly Listed	Not Listed	Directly Listed			

Legend:
Directly Listed
Indirectly Listed
Not Listed

Source: PMI, 2013
Numbers refer to corresponding chapter in *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)* – Fifth Edition.

explicitly listed as a tool/technique for any of the three Project Quality Management Knowledge Areas.

Across all 10 Knowledge Areas, expert judgment is listed as a tool/technique *five times* more frequently than the next most commonly listed project management tool/technique. By considering this fact, we may conclude that expert judgment plays an important role in project management.

Despite its prevalence as a project management tool/technique, expert judgment lacks full description within the *PMBOK® Guide*, which provides the following short definition:

> Judgment provided based upon expertise in an application area, knowledge area, discipline, industry, etc. as appropriate for the activity performed. Such expertise may be provided by any group or individual with specialized education, knowledge, experience, skill, or training. (PMI, 2013, p. 538)

This definition is expanded upon in only a few of the project management processes, but any additional detail is typically confined to minor descriptions of who might be included as experts, lists of quantities or qualities to be characterized using expert judgment, or precautions about taking expert bias into account. There is no description about how the tool/technique of expert judgment may or should be applied. As a result, it would be difficult for a practitioner to understand exactly how to apply expert judgment as a tool/technique using only the definition provided.

By comparison, tools/techniques such as the critical path method (CPM) and the probability-impact (P-I) matrix, which are referenced in only a single specific project management process, are much more fully described in the *PMBOK® Guide*, such that a practitioner would be able to more easily apply that particular tool/technique. Additionally, although these two less frequently invoked tools/techniques (i.e., CPM and P-I matrix) are contained in the *PMI Lexicon of Project Management Terms* (2012), the ubiquitous expert judgment is not.

There is a vast amount of literature about expert judgment elicitation. There are even entire books devoted to the subject of expert judgment elicitation (e.g., Ayyub, 2001; Cooke, 1999; Meyer & Booker, 2001; O'Hagan et al., 2006). Yet, none of these books focuses exclusively on the expert judgment needs of project managers. In project management, expert judgment includes both qualitative and quantitative methods, both direct and indirect elicitation, and both individual and consensus aggregation. Examples of project management expert judgment elicitation methods include brainstorming, the Delphi method (Dalkey & Helmer, 1963), direct point elicitation, distribution estimation (including the prominent PERT [program evaluation and review technique], a three-point estimation technique developed by Malcolm, Roseboom, Clark, and Fazar [1959]), the analytic hierarchy process developed by Saaty (1980), and scaling methods (e.g., Kent, 1964). Though many of these expert judgment elicitation techniques are well established, much has been learned over the past several decades about how they can be improved (e.g., Armstrong, 2011). Yet, despite the fact that the number of articles about expert judgment has increased steadily in recent years (Jørgensen & Shepperd, 2007), extensive study about how to improve expert judgment in project management seems to be confined to two specific aspects of project management: time and cost estimation (which represent only a small portion of the project management processes that call on expert judgment as a tool/technique).

Even with considerable advances in the areas of time and cost estimation, it seems they have not been incorporated into the most frequently used practitioner references (including the *PMBOK® Guide*), the basic software packages (such as Microsoft Project), or the project management texts (e.g., Kerzner, 2012; Mantel, Meredith, Shafer, & Sutton, 2010). Therefore, based upon a preliminary and limited review of the relevant literature (before embarking on this study), even though much empirical and theoretical work has been conducted regarding the elicitation of expert judgment, we anticipate that these developments have not been widely adopted into the practice of project management.

1.2 Problem Statement

This lack of definition for expert judgment represents a significant gap in the *PMBOK® Guide* toolkit. Without well-designed elicitation processes, expert judgment is subject to known flaws that can render the resulting estimates inaccurate. When project management processes are based upon flawed judgments and estimates, projects are susceptible to missed deadlines, budget overruns, and/or failure to meet stakeholder expectations. This is not uncommon. In general, project management practice lags behind theory, as described by Ahlemann, Arbi, Kaiser, and Heck (2013):

> Despite [a] long tradition of prescriptive research, project management methods suffer a number of problems, such as lack of acceptance in practice, limited effectiveness, and unclear application scenarios. We identify a lack of empirical and theoretical foundations as one cause of these deficiencies. (p. 44)

Expert judgment suffers from all three of the above-noted problems. Whether it is known or not, the prescriptive research about how to improve expert judgment has not been widely adopted into project management practice. Further, given that there is an over-reliance on ad hoc and qualitative methods, elicitation of expert judgment is less effective than it might be. Additionally, with such a wide range of diverse application scenarios (e.g., see Table 1), it is not apparent that the most appropriate methods are being applied for each project management process and scenario.

Thus, it is critical that project managers have access to a foundational set of guidelines in order to handle expert judgment appropriately and make more accurate project estimates. This research project seeks to narrow the *PMBOK® Guide* toolkit gap by addressing the following question: How might expert judgment be better defined to ensure that the most accurate information is elicited for use in today's project management processes?

1.3 Objectives and Scope

In order for this *translational* research to address that basic research question, the following three specific aims were identified:

1. Identify the state of the art/science in expert judgment elicitation broadly (e.g., across a variety of disciplinary areas such as engineering, political science, and environmental management).
2. Determine the state of the practice in expert judgment elicitation for project management.
3. Narrow the theory-practice gap in expert judgment elicitation for project management.

1.4 Methodology

The research contained in this report involves a mixed methodology (Creswell, 2013) to achieve the overarching goal of developing a *practitioner-ready* expert judgment reference for project managers.

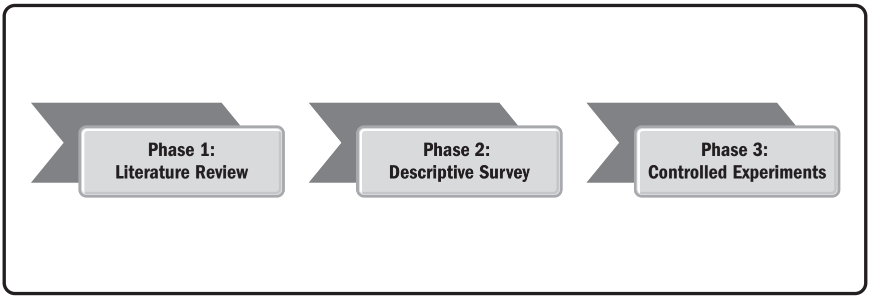


Figure 1 Research design.

As illustrated in Figure 1, this mixed methodology comprises three interrelated, sequential phases that address the three specific aims of the research:

- Phase 1 comprised a *literature review* to identify the state of the art/science in expert judgment elicitation broadly.
- Phase 2 employed a *descriptive survey* to determine the state of the practice in expert judgment elicitation in project management.
- After identifying several of the theory-practice gaps, phase 3 used *controlled experiments* to determine the effectiveness of different methods for selecting experts.

Because the design types noted above (i.e., *literature review, descriptive survey*, and *controlled experiments*) are straightforward and well established, the details of those design types will not be provided here. Rather, for each of the three phases, details regarding type of research, rationale, process, type of data, source and selection of data, expected outcomes, and potential problems will be provided in the following sections. Further, the general design type has been identified because research design is logical, rather than logistical, in nature (Yin, 2003).

The following is a high-level description of some of the key elements of the methodological procedures for each of the three phases of this research:

1.4.1 Phase 1

- *Goal:* Identify the state of the art/science in expert judgment elicitation broadly.
- *Design Type:* Literature Review
- *Rationale:* In order to determine which expert judgment techniques are most appropriate for today's project management processes, it was necessary to first determine the spectrum of expert judgment techniques available for application. A review was conducted of the literature from across many disciplinary areas (e.g., engineering, political science, and environmental management) in addition to project management.
- *Process:* A systematic literature review process (Brereton, Kitchenham, Budgen, Turner, & Khalil, 2007) was employed.
- *Type of Data:* Qualitative
- *Data Source and Selection:* Data were taken from a systematic search of selected bibliographic databases containing published research studies.
- *Expected Outcomes:* It was anticipated that hundreds of studies and dozens of expert judgment techniques would be identified from the many disciplinary areas. Many of the expert judgment techniques are not widely used in project management. These will provide opportunities to enhance the accuracy of the estimates conducted in project management.
- *Potential Problems/Alternative Approaches:* Expert judgment of some form is used in virtually all disciplinary areas, so one major challenge was to develop an effective means of narrowing the search parameters while maintaining an adequate pool of literature. Also, because available subscriptions to certain

desired databases are limited, some articles were not readily obtainable. In those cases, alternative sources (such as interlibrary loans) were sought to obtain the relevant publications.

1.4.2 Phase 2

- *Goal:* Determine the state of the practice in expert judgment elicitation in project management.
- *Design Type:* Descriptive Survey
- *Rationale:* It was anticipated that a variety of expert judgment elicitation techniques are in use in project management today. Thus, it was essential to identify prominent practices in an effort to identify current practices.
- *Process:* A standard survey methodology (Groves et al., 2013) was observed.
- *Type of Data:* Qualitative and quantitative
- *Data Collection:* An online survey was administered to project management professionals using the Project Management Institute Survey Links program. Based on previous similar studies, it was anticipated that there would be roughly 400 to 500 respondents. Demographic variables included job position, experience, field of specialty, and office location. Expert judgment study variables included frequency of use, purpose of use, context, policies, and methods.
- *Expected Outcomes:* It was anticipated that the response rate would be sufficient to provide meaningful results because the survey would be relatively short, relevant to project management professionals, and administered through Survey Links.
- *Potential Problems/Alternative Approaches:* A low response rate to the online survey was anticipated to be a potentially significant problem. In such a case, an alternative would have been to deploy the survey

through PMI's regional chapters, as was done in a recent earned value management study (Song, 2010).

1.4.3 Phase 3

- *Goal:* Identify how general expert judgment methods (e.g., expert selection) can be adapted to project management.
- *Design Type:* Controlled Experiment
- *Rationale:* Once a collection of potential practices (from phase 1) and current practices (from phase 2) of expert judgment in project management was identified, it was necessary to determine means of identifying experts to be used for judgment elicitation.
- *Process:* An expert elicitation protocol was employed. The protocol consisted of two main parts: expert training and expert elicitation. Subjects were asked to make estimations about known quantities in a variety of modes.
- *Type of Data:* Quantitative and qualitative
- *Data Collection:* Actual estimations by project management professionals were to be collected using a standard expert elicitation protocol at professional society meetings (such as the PMI® Global Congress—North America, or PMI regional chapter meetings) and also virtually using elicitation techniques such as the Delphi method.
- *Expected Outcomes:* It was anticipated that several rounds of experiments would need to be successfully conducted to provide sufficient results.
- *Potential Problems/Alternative Approaches:* It was anticipated that gaining agenda time at the various Project Management Institute events to conduct the experiments would be difficult. In such a case, voluntary participation would be sought outside of the agenda or via online methods (e.g., through webinar format). Alternative professional and academic venues were also sought.

1.5 Data Analysis

Phase 1 involved a *literature review* that was designed to identify the state of the art/science in expert judgment elicitation broadly. In this portion of the study, the data were the individual articles and the findings of the studies contained therein. Each article was recorded with specific attention being paid to the expert judgment elicitation methods used. This phase did not attempt to complete a meta-analysis of the studies' results.

Phase 2 involved a *descriptive survey* designed to determine the state of the practice in expert judgment elicitation in project management. The data were the responses of the survey participants. Two forms of analysis were completed in this portion of the study. First, summary statistics were developed for all closed responses. Then, a small cross-sectional analysis was conducted using the demographics as the independent variables to determine if certain expert judgment methods were used in certain situations. For the open responses, contextual content analysis was conducted using raters and software (Krippendorff, 2012).

Finally, after several theory-practice gaps were identified in the first two phases, phase 3 employed *controlled experiments* to address the gap in how to select experts to provide judgments for project management. The data were the participants' responses on the expert elicitation worksheet, as well as their scores on the critical thinking, fluency, and numeracy instruments. By gathering experts' estimation of quantities on the elicitation forms, processes of expert selection were evaluated. Most often, the expert judgments consisted of a series of estimates or value judgments to be analyzed as quantities. For example, in the case of elicited distribution parameters, responses will be treated using an arcsine transformation for ease of comparison. Hit rates were established to determine the accuracy of experts, and standard statistical methods were used to determine the most effective expert judgment elicitation methods.

1.6 Organization of Report

The report will follow an organization that is aligned with the design of the research. This current chapter provides an overview of the study. Chapter 2 provides an overview of the state of the art/science of expert judgment broadly across many domains and disciplines. This state-of-the-art/science information was obtained through a review of the literature in phase 1 of the research project. Chapter 3 provides a summary of the state of the practice for using expert judgment in project management. This state-of-the-practice information was obtained through a survey of project management practitioners in phase 2 of the research project. Chapter 4 provides information about expert selection. This practical information was obtained through a pair of experiments developed to test two methods for selecting experts (i.e., phase 3 of the research project). Chapter 5, the final chapter of this report, provides a discussion about how the information gathering in this research project can be used by project management practitioners to improve their use of expert judgment as a tool/technique.

CHAPTER 2

State of the Art/ Science

The first phase of this study involved determining the state of the art/science broadly. Rather than focusing solely on project management and what "should be," this phase focused on a wide range of disciplinary areas (such as engineering, political science, environmental management, and medicine) to explore what "might be" in regard to the practice of eliciting expert judgment.

2.1 Theory-Practice Gap

The idea that a gap or divide exists between theory and practice has been widely discussed (e.g., Brendillet, Tywoniak, & Dwivedula, 2015; Kraaijenbrink, 2010; Sandberg & Tsoukas, 2011; Van de Ven & Johnson, 2006). Additionally, several studies suggested that such a theory-practice gap exists within project management as well (e.g., Koskela & Howell, 2002; Söderlund, 2004; Svejvig & Andersen, 2015).

This study clearly shows that a gap does indeed remain between the theory and practice of expert judgment within project management. The first phase of this research project, reported in this chapter, identifies a body of the most relevant theory. The second phase of the research project, reported in the next chapter, identifies the current practice of expert judgment in project management. Comparing the two reveals the theory-practice gap that exists.

Although many expert judgment elicitation techniques are well established, much has been learned over the past several decades about how they can be improved (e.g., Armstrong, 2011). Additionally, even though the study of expert judgment within the context of project management has been steadily increasing (Jørgensen & Shepperd, 2007), that study does not span the breadth of the discipline.

Much advancement has been made in the field of expert judgment with regard to project time estimation (Trendowicz, Munch, & Jeffery, 2011). The most common form of time estimation is the program evaluation and review technique (PERT) a three-point estimation technique (developed in 1959 by Malcolm et al.). The original PERT can be found in virtually any textbook on project management. More recently, many advances have been made in the PERT, including new and improved expressions of PERT mean and variance (e.g., Golenko-Ginzburg, 1989; Hahn, 2008; Herrerías, García, & Cruz, 2003; Herrerías-Velasco et al., 2011), alternative distributional forms (e.g., Garcia, Garcia-Perez, & Sanchez-Granero, 2012; Herrerías-Velasco et al., 2011; Premachandra, 2001), and new ways to estimate key parameters (e.g., Sasieni, 1986; van Dorp, 2012). Yet, these advances have not been incorporated into either the academic resources (e.g., texts) or the professional practitioner resources (e.g., software, guides). As a result, even in one of the most widely studied areas of project management expert judgment, a gap remains between the theory and the practice.

Another area of advancement and innovation in expert judgment has been observed in software project management—specifically in the area of cost estimation. Because estimates often propagate throughout an entire project plan (Sudhakar, 2013), and since flawed estimation has been identified as one of the top failure factors in software project management (Dwivedi et al., 2013), cost estimation has been identified as critical to project success. For example, in a study of 250 complex software projects (Jones, 2004), less than 10% of the projects were successful (i.e., less than six months over schedule, less than 15% over budget). In a continuing effort to improve cost estimation, a variety of new and improved methods has been identified (e.g., Kim & Reinschmidt, 2011; Li, Xie, & Goh, 2009; Liu & Napier, 2010). Just as with time estimation, these new approaches have not been widely adopted (Trendowicz, Munch, & Jeffery, 2011). Again, even though the theory has advanced, there is a gap in that the practice lags behind theoretical improvements.

Additionally, regardless of which expert judgment elicitation methods are used in project management, estimation is known to be flawed (Budzier & Flyvberg, 2013; Flyvberg, 2006; Flyvberg, Holm, & Buhl, 2005). Noted in the seminal work of Kahneman, Slovic, and Tversky (1982) and expanded through an abundance of recent research (as summarized in Lawrence, Goodwin, O'Connor, & Onkal, 2006), expert judgments are subject to well-known cognitive biases. One of the most common forms of cognitive bias in experts is overconfidence (Lichtenstein, Fischhoff, & Phillips, 1981; Lin & Bier, 2008). There have been many studies to improve how we elicit expert judgment to reduce overconfidence and, in turn, increase accuracy by changing the mode of elicitation (e.g., Soll & Klayman, 2004; Soll & Larrick, 2009; Speirs-Bridge et al., 2010; Teigen & Jorgensen, 2005; Welsh, Lee, & Begg, 2008, 2009; Winman, Hansson, & Juslin, 2004), by including feedback (e.g., Bolger & Önkal-Atay, 2004; Haran, Moore, & Morewedge, 2010; Herzog & Hertwig, 2009; Rauhut & Lorenz, 2010; Vul & Pashler, 2008), and through other means. Here, too, the theoretical enhancements to mitigate the adverse impacts of overconfidence have not been widely adopted into practice.

2.2 Method

A 10-step review process (Brereton et al., 2007) was used to conduct this literature review. Figure 2 provides an overview of the 10 steps involved in this particular process. These steps could be aggregated into three main stages (i.e., plan review, conduct review, and document review).

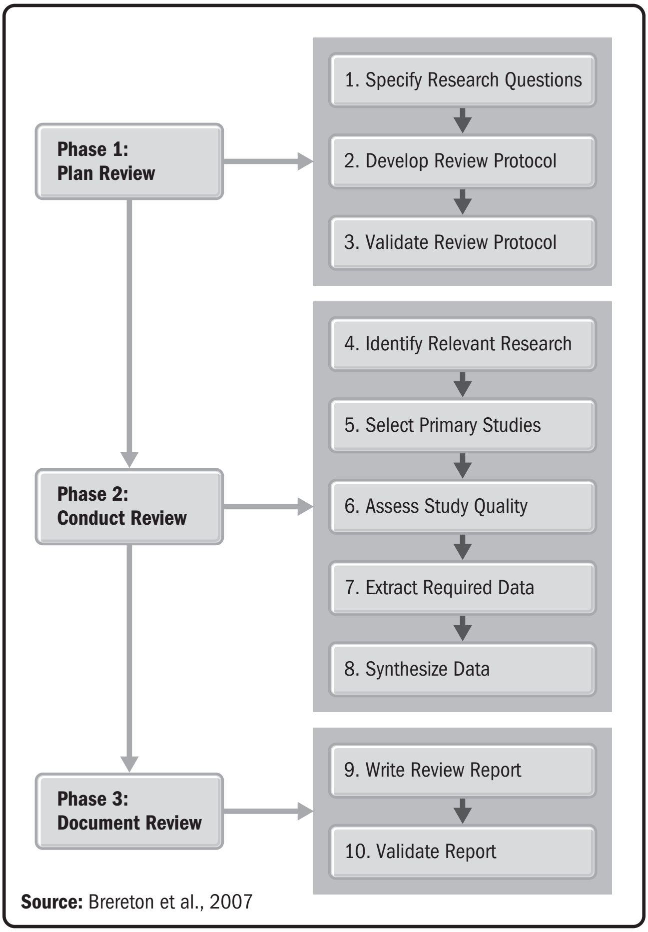


Figure 2 Literature review process.

2.3 Data and Sample

Because the most recent compendium of expert judgment elicitation (i.e., O'Hagan et al., 2006) published the results of a comprehensive literature review completed in 2005, this literature review set out to emulate the process of that work by examining the most recent decade (which would have brought the results of that work up to the present). However, it was quickly determined that such an expansive review of the most recent decade of

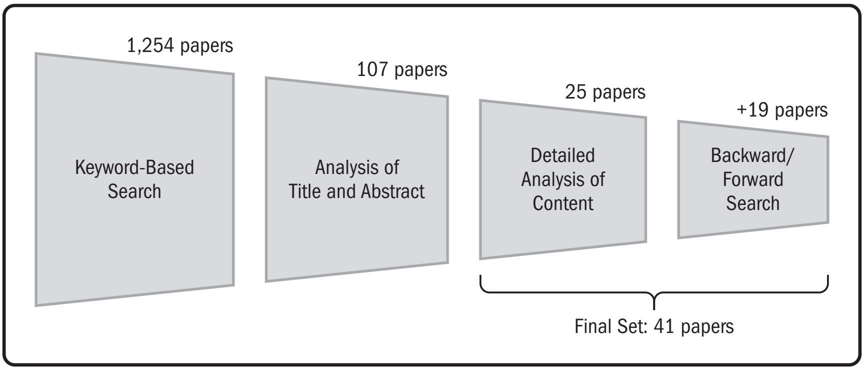


Figure 3 Summary of literature review "funnel."

literature (using similar search terms, parameters, and databases) was not possible. The BEEP (Bayesian Elicitation of Experts' Probabilities) project commissioned by the UK National Health Service (as reported by O'Hagan et al., 2006), which reviewed more than two decades' worth of literature, was conducted by a large team of researchers over a multiyear period. Because our study was conducted by a small team of researchers over a few months, such a comprehensive decade study period proved to be beyond the scope of the study. Instead, this study focused on the most recent year (i.e., mid-2013 to mid-2014) of the decade since the BEEP project was completed. This compromise was deemed adequate because additional sources of information would be uncovered in the forward and backward searches of the relevant literature from that one-year period. Figure 3 summarizes the results of the literature review conducted in this study.

Similar to the BEEP project (as reported by O'Hagan et al., 2006), this study searched the ISI Science, Social Sciences, and Humanities Citation Indices under the terms *expert judgment*, *expert opinion*, and *elicitation* for the most recent one-year period. More than 1,200 articles were identified as relevant and investigated further. A careful reading of the abstracts of those references led to the selection of more than 100 sources, whose full text was retrieved and read. The resulting detailed content analysis of these articles yielded 25 papers that were relevant to the topic of expert judgment elicitation (for project management). In an attempt to address the entire decade-long period since the BEEP project, a forward and backward search was conducted and an additional 19 articles were identified as relevant, bringing the final set of articles to 41 articles. In comparison, the BEEP project identified 13,000 references from keyword searches. The most relevant 2,000 were narrowed down to 400 based upon a review of the abstracts. The remaining 400 were read in detail. As is typical of this type of work, though the literature review was intended to be comprehensive in scope, there will inevitably be omissions. Further, the discussion and emphases contained in this report about those references (and the attempt to translate the ideas from many varied disciplines to the world of project management) will reflect the perspective of the author.

2.4 Analysis and Results

In order to provide this review of the literature structure, the first step was to identify a framework by which the results could be organized. To accomplish this, several general (as well as some specific) expert judgment elicitation processes and protocols were examined. There was a wide variety of protocols that involved a few or several steps. For example, the U.S. Environmental Protection Agency (2011) has as few as three steps; Catenacci, Bosetti, Fiorese, and Verdolini (2015) offer a three-phase protocol; Meyer and Booker (2001) suggest seven steps; Ayyub (2001) provides eight steps; Aliakabargolkar and Crawley (2014) offer a 10-step model for space exploration; and the EU Atomic Energy Community protocol (Cooke & Goossens, 2000) has as many as 15 steps. Table 2 examines a few of the most prominent processes in chronological order. It starts with the protocol designed in the seminal work of the U.S. Nuclear Regulatory Commission (NUREG) (leftmost column). It then proceeds to a protocol developed for the EU Atomic Energy Community by the researchers from the Technical University in Delft, the Netherlands (second column). Their protocol

Table 2 Comparison of some prominent expert judgment processes.

	NUREG-1150 (Comer et al., 1984)	EUR 18820 (Cooke & Goossens, 2000)	Expert Opinion (Aayub, 2001)	Practical Guide (Meyer & Booker, 2001)	Uncertain Judgments (O'Hagan et al., 2006)
Frame the Problem (INITIATE)	1. Selection of Issues	1. Definition of Case Structure 2. Identification of Target Variables 3. Identification of Query Variables 4. Identification of Performance Variables	1. Identify Need of an Expert Elicitation Process 3. Define Study Level 5. Identify and Select Technical Issues	1. Selecting the Question Areas and Particular Questions 2. Refining the Questions	1. Background and Preparation
Plan the Elicitation (PLAN)		7. Definition of Elicitation Format Document 8. Dry Run Exercise	2. Select Study Leader 4. Select Technical Integrator and Facilitator	4. Selecting the Components of Elicitation 5. Designing and Tailoring the Components of Elicitation to Fit the Application 6. Practicing Elicitation and Training In-House Personnel	4. Structuring and Decomposition
Select the Experts (EXECUTE)	5. Selection of Experts	5. Identification of Experts 6. Selection of Experts	6. Identify and Select Experts and Peer Reviewers	3. Selecting and Motivating the Experts	2. Identify and Recruit Experts
Train the Experts (EXECUTE)	3. Training in Elicitation Methods	9. Expert Training Session	7. Discuss and Refine the Issues 8. Train the Experts for Elicitation		3. Motivating and Training the Expert(s)
Elicit Judgments (EXECUTE)	4. Presentation and Review Issues 5. Preparation of Expert Analyses 6. Expert Review and Discussion 7. Elicitation of Experts	10. Expert Elicitation Session	9. Facilitate Group Interactions and Expert Opinions	7. Eliciting and Documenting Expert Judgments	5. The Elicitation
Analyze and Aggregate Judgments (MONITOR AND CONTROL)	6. Composition and Aggregation of Judgments 7. Review by Experts	11. Combination of Expert Assessments 12. Discrepancy and Robustness Analysis 13. Feedback	10. Analysis, Aggregation, Revisions, Resolution of Disagreement, and Consensus Estimation of Needed Quantiles 11. Administer Peer Review		
Document Results (CLOSE)		12. Post-Processing Analyses 13. Documentation	12. Document Process and Communicate Results		

built on the work of the NUREG-1150 and is also based on their experience completing hundreds of studies and compiling thousands of judgments (Cooke & Goossens, 2008). Next, the generic protocols are presented from a trio of books on expert judgment (opinion) (columns three through five). These references have been widely used by practitioners conducting expert judgment elicitation. It should be observed that there are many similarities and few differences among the five elicitation protocols presented. Some offer peer review, some have much more detailed planning, and some have more interactive elicitation methods, but all follow a similar pattern that can be organized into the five Project Management Process Groups of the *PMBOK® Guide*: initiation, planning, executing, monitoring and controlling, and closing.

In the previous table, note that the numbering provided corresponds to that of the original source. In some cases, the ordering of the numbered process steps is out of sequence of the original source to demonstrate the relationship to the generic seven steps used in this study (i.e., those provided in the leftmost column). In an effort to integrate the various protocols of elicitation from Table 2, the following generic seven-step protocol is proposed (as presented in the row headers on the left side of the table):

1. Frame the Problem
2. Plan the Elicitation
3. Select the Experts
4. Train the Experts
5. Elicit Judgments
6. Analyze/Aggregate Judgments
7. Document/Communicate Results

Note that this summary protocol is nicely aligned with the phases of project management from the *PMBOK® Guide*. The summary protocol includes an initiation phase (step 1), a planning phase (step 2), an execution phase (steps 3–5), a monitor/control phase (step 6), and a closeout phase (step 7). Such an orientation would be well understood by project management practitioners.

Two prominent compilations of effective practices in expert elicitation have been overlaid upon the suggested seven-step protocol to demonstrate its coherence. Table 3 shows 12 principles of expert judgment recommended by Jørgensen (2004) and 10 recommendations suggested by Kynn (2008) that are based

Table 3 Compilation of ways to improve expert judgment.

	12 Principles (Jørgensen, 2004)	10 Recommendations (Kynn, 2008)
Frame the Problem		
Plan the Elicitation	• Avoid conflicting estimation goals • Avoid irrelevant and unreliable estimation information • Use documented data from previous development tasks • Estimate top-down and bottom-up independently of each other	• Only ask questions from within the area of expertise by using familiar measurement • Decompose the elicitation into tasks that are as "small" and distinct as possible • Be specific with wording (use frequency representation where possible with an explicit reference class)
Select the Experts	• Find experts with relevant domain background and good estimation records	
Train the Experts	• Provide estimation training opportunities	• Familiarizing experts with the elicitation process is beneficial, but training questions are only effective for calibration when directly related to the test questions • Scoring rules can be used as a training device, but they need to be transparent • A brief review of probability concepts may be helpful
Elicit the Judgments	• Use estimation checklists • Assess the uncertainty of the estimate • Ask the experts to justify and criticize the estimates • Provide feedback on estimation accuracy and development task relation • Evaluate estimation accuracy, but avoid high evaluation pressure	• Do not lead the expert by providing sample numbers upon which the expert may anchor • Ask the expert to discuss the estimates, giving evidence both for and against • Offer process feedback about the task and the probability assessments; give experts summaries of estimates and allow reconsideration of estimates
Aggregate/Analyze the Judgments	• Combine estimates from different experts and estimation strategies	• If possible, duplicate the elicitation procedure with the same experts at a later date to check self-consistency of experts
Document/Communicate the Results		

upon separate, comprehensive reviews of the literature. These provide a preview of some of the findings about ways to improve expert elicitation (and would serve as a good starting point to learn more about how to improve expert judgment).

With the exception of the first and last steps, the suggestions contained in Table 3 span the generic seven-step protocol of expert judgment elicitation proposed. Regarding a protocol of expert judgment elicitation, all seven steps should be retained for completeness until they can be properly tested (which was beyond the scope of this study). Meantime, for simplicity and consistency, the five middle steps provided in Table 3 will become the organizing frame of the literature review to identify the state of the art/science as it pertains to expert judgment elicitation.

2.4.1 Planning the Elicitation

Once the problem has been framed (and the desired data and information to be elicited have been identified), planning commences and an appropriate method must be chosen to elicit the requisite expert judgment.

In a broad sense, the form of the data and information may be considered either qualitative or quantitative using Stevens's (1946) scales of measurement—nominal and ordinal data being primarily qualitative; interval and ratio data being predominantly quantitative data (also referred to as "weak" and "strong" data scales, respectively, by Wachowicz and Blaszczyk [2013]). It has been suggested that some experts have a preference for one form of elicitation over the other (i.e., quantitative versus qualitative) (Larichev & Brown, 2000). This dichotomy has also been tested to demonstrate how experts' numeracy or fluency will affect their ability to provide judgments about quantitative or qualitative information, respectively (Fasolo & Bana e Costa, 2014).

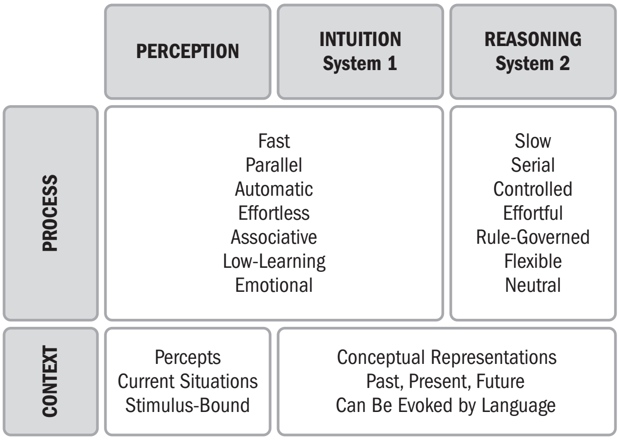
Additionally, recent neuroscience has further established this dichotomy of judgment types (i.e., qualitative and quantitative) by examining how the expert's brain functions when rendering these two types of judgments. In recent decades, there has been a growth in our understanding of the human brain and its functioning (across a wide variety of contexts) as a result of advances in brain imaging using functional magnetic resonance imaging (fMRI). One study identified that the task-positive network (TPN) regions of the brain have been shown to be activated in a broad range of attention-focused tasks (e.g., Buckner, Andrews-Hanna, & Schacter, 2008; Fox, Corbetta, Snyder, & Vincent, 2006). Because of this, the TPN would likely be the area of the brain activated through *evaluative* expert judgment situations. On the other hand, the default mode network (DMN) regions of the brain have been shown to be activated in idea generation (e.g., Beaty et al., 2014; Kleibeuker, Koolschijn, Jolles, de Dreu, & Crone, 2013), envisioning the future (e.g., Uddin, Kelly, Biswal, Castellanos, & Milham, 2009), and creativity or insightful problem solving (e.g., Subramaniam, Kounios, Parrish, & Jung-Beeman, 2009; Takeuchi et al., 2011). Therefore, in *generative* expert judgment situations, the DMN would likely be activated. Interestingly, activities in the TPN tend to inhibit activities in the DMN, and vice versa (e.g., Boyatzis, Rochford, & Jack, 2014; Jack, Dawson, & Norr, 2013). Therefore, such evidence from neuroscience would seem to emphasize the importance of matching methods for eliciting expert judgment to the desired form of information. So, identifying whether qualitative or quantitative information is needed will determine whether *generative* or *evaluative* methods are needed.

Therefore, in the context of project management, two basic forms of expert judgment elicitation methods may be suggested: *generative* and *evaluative*. On one hand, *generative* elicitation methods would yield a list of *generated* items, scenarios, lists, and so forth. For example, in the Collect Requirements process (identified as 5.2 in the *PMBOK® Guide*), a *generative* elicitation process would be used to generate a list of requirements using stakeholder input. On the other hand, the *evaluative* elicitation methods would be used to *evaluate* (or quantify) a specific phenomenon of interest. In the Estimate Activity Durations process (identified as 6.5 in the *PMBOK® Guide*), any one of many *evaluative* expert judgment elicitation processes could be employed to create the requisite time estimates.

In order to be beneficial to project management practitioners, all *PMBOK® Guide* project management processes (which specifically list expert judgment as a tool or technique—see Table 1) will be categorized as either *generative* or *evaluative* in Table 4. This set of processes was determined to be either generative or evaluative—the two suggested basic forms of expert judgment elicitation for project management. If the primary output(s) of a process is/are numerical in nature (e.g., cost, time, probability estimates), the process is deemed to be *evaluative*. If the primary output(s) of a process is/are verbal in nature

Table 4 Categorization of *PMBOK® Guide* processes.

Generative PMBOK® processes	Process outputs
4.1 Develop project charter	Project charter
4.2 Develop project management plan	Project management plan
4.3 Monitor and control project work	Deliverables, work performance data, change requests, updates
4.4 Perform integrated change control	Change requests, work performance reports, updates
4.5 Close project or phase	Approved change requests, change log, updates
5.1 Plan scope management	Scope management plan, requirements management plan
5.4 Create work breakdown structure	Scope baseline, updates
6.1 Plan schedule management	Schedule management plan
6.2 Define activities	Activity list, activity attributes, milestone list
6.4 Estimate activity resources	Activity resource requirements, resource breakdown structure, updates
7.1 Plan cost management	Cost management plan
9.1 Plan human resource management	Human resource management plan
10.3 Control communications	Work performance information, change requests, updates
11.1 Plan risk management	Risk management plan
11.2 Identify risks	Risk register
12.1 Plan procurement management	Procurement management plan, SOW, documents, change requests
12.2 Contract procurements	Selected sellers, agreements, resource calendars, change requests, updates
13.4 Control stakeholder engagement	Work performance information, change requests, updates
Evaluative PMBOK® processes	**Process outputs**
6.5 Estimate activity durations	Activity duration estimates, updates
7.2 Estimate costs	Activity cost estimates, basis of estimates, updates
7.3 Determine budget	Cost baseline, project funding requirements, updates
11.3 Perform qualitative risk analysis	Project management plan updates (e.g., P-I matrix)
11.4 Perform quantitative risk analysis	Project management plan updates (e.g., probabilistic information)
11.5 Plan risk responses	Project management plan updates (e.g., risk register)



Source: Kahneman, 2011, p. 1451

Figure 4 Cognitive systems.

(e.g., lists, plans, registers), then the process is deemed to be *generative* in nature.

To further reinforce the science behind expert judgment, let's now look at a framework for classifying how judgments are made (Kahneman, 2011). In Figure 4, there are two "systems" that describe different ways in which the mind thinks. In describing these cognitive systems, Kahneman (2011) invokes agency upon the "systems" as a way to better describe the differences in how they work, even though they are not systems per se. Stanovich and West (2000) use the more neutral term *type* in their description. System 1 relies on association and produces impressions about the attributes of objects from perception and thought. It is automatic and quick-acting. The judgments from system 1 are always explicit and intentional, whether or not they are expressed. System 2, on the other hand, is what we normally think of when we consider judgment; it is deliberate and conscious.

Reconsidering the two forms of expert judgment, it would seem natural that system 1 would do a good job of *generative* elicitation and system 2 would be best used in *evaluative* elicitation.

Table 5 Generative expert judgment elicitation methods.

Appreciative Inquiry
Brainstorming
Brainwriting
Clustering
Codiscovery (Barnum, 2010)
Delphi Technique/Method (Dalkey & Helmer, 1963)
Dual Verbal Elicitation (McDonald, Zhao, & Edwards, 2013)
Metaphors (e.g., Jacobs, Oliver, & Heracleous, 2013; Cornelissen, 2005)
Nominal Group Technique (Delbecq & Van de Ven, 1975)
Photo Narrative (Parke et al., 2013)
Scenario Planning
Think-Aloud Protocols

Because certain elicitation methods are better suited for yielding qualitative information and others are more suited for providing quantitative information, once it has been determined which of the basic forms of expert judgment (i.e., generative or evaluative) is needed, an appropriate elicitation method must be selected. Based upon a review of the literature, expert judgment elicitation methods have been labeled either generative or evaluative. Table 5 provides a list of some of the generative expert elicitation methods that may be chosen.

Similarly, Table 6 provides a list of some of the evaluative expert elicitation methods that may be chosen.

Because there is a wide variety of methods by which to elicit expert judgments (e.g., those listed in Tables 5 and 6), the choice of elicitation method will depend upon both the type of information needed and the type of expertise available. After determining the form of the expert judgment needed for a particular project management process, it will be necessary to determine the type of expertise that is necessary (and determine if it is available).

2.4.2 Selecting the Experts

There are many considerations to take into account when selecting experts. First, it is important to identify the requisite expertise

Table 6 Evaluative expert judgment elicitation methods.

<table>
<tr><th colspan="2">Point Distribution Methods</th></tr>
<tr><td rowspan="2">Distribution Estimation Methods
• One-point estimation (e.g., 80% sure between __ and __)
• Two-point estimation (e.g., 90% sure greater than __ and 90% sure less than __)
• Three-point estimation (e.g., 5%, 50%, 95%, which includes PERT 3-point estimation, Dalkey & Helmer, [1963])
• Four-point estimation (e.g., min, 50%, max, and participant-assigned Confidence Intervals) (Speirs-Bridge et al., 2010)
• Six complementary intervals (as described in Grigore, Peters, Hyde, & Stein, 2013)
• SPIES (i.e., subjective probability interval estimates) (e.g., assigns frequencies to pre-assigned bins) (Haran et al., 2010)
• MOLE (i.e., more or less estimation) (Welsh et al., 2008, 2009)
• Graphical allocation (e.g., assigns 100 “chips” to a set of potential options) (Catenacci, Verdolini, Bosetti, & Fiorese, 2013)</td><td>Paired Comparison Methods
• Bradley-Terry model (1952)
• Analytic hierarchy process (Saaty, 1980)
• Bayesian methods (e.g., Merrick, van Dorp, & Singh, 2005; Szwed, van Dorp, Merrick, & Singh, 2006)
• Interactive slider method (Curtis & Wood, 2004)
• Negative exponential life (NEL) Model
• Rank order centroid technique (Barron, 1992)
• Thurstone Model (1927)</td></tr>
<tr><td>Chance Methods
• Odds ratios
• Lottery wheels
• BRET (i.e., bomb risk elicitation task) (Crosetto & Filippin, 2013)
• Ordered lottery choice (Eckel & Grossman, 2002)</td></tr>
<tr><td>Scaling Methods (e.g., Torgersen, 1958)
• Continuous rating
• Discrete rating (e.g., Kent, 1964)
• Order ranking (e.g., Labovitz, 1970)</td><td>Other
• Delphi technique/method (Dalkey & Helmer, 1963)
• MACBETH (measuring attractiveness by a categorical-based evaluation technique) (Bane e Costa, De Corte, & Vansnick, 2012)
• Fuzzy logic approaches
• Reference class forecasting (Flyvberg, Holm, & Buhl, 2002)</td></tr>
</table>

that will be required to accomplish the process or task at hand. There are several compelling definitions of expertise. For example, Collins and Evans (2007) offer a taxonomy of expertise, as shown in Table 7.

Woods and Ford (1993) describe four fundamental ways in which expertise (as opposed to amateur or lay judgment) is demonstrated:

- Expert knowledge is grounded in specific cases.
- Experts represent problems in terms of formal principles.
- Experts solve problems using known strategies.
- Experts relay less on declarative knowledge and more on procedural knowledge.

Table 7 Taxonomy of expertise.

Type	Characteristics
Contributory expertise	Fully developed and internalized skills and knowledge, including an ability to contribute new knowledge and/or teach
Interactional expertise	Knowledge gained from learning the language of specialist groups, without necessarily obtaining practical competence
Primary source knowledge	Knowledge from primary literature includes basic technical competence
Popular understanding	Knowledge from media, with little detail, less complexity
Specific instruction	Formulaic, rule-based knowledge, typically simple, context-specific and local

Source: Collins & Evans, 2007

Shanteau (1992) argues that evidence of relevant experience and training includes the following:

- Certifications such as academic degrees or professional training
- Professional reputation of the expert (as a potentially reliable guide)
- Impartiality
- Multiplicity of viewpoints (i.e., consideration of multiple forms of data and perspectives

Hora and von Winterfeldt (1997) suggest the following criteria for scrutinizing experts (particularly in a highly public and controversial context):

- Tangible evidence of expertise
- Reputation
- Availability and willingness to participate
- Understanding of the general problem area
- Impartiality
- Lack of an economic or personal stake in potential findings

Experts selected for the elicitation should possess high professional standing and widely recognized competence (Burgman, McBride, et al., 2011). The group of experts should represent a diversity of technical perspectives on the issue of concern. Although experience does not necessarily yield expertise, there is some evidence to suggest that professionals with more expertise are subject to less bias and better judgments (Adelman & Bresnick, 1992; Adelman, Tollcott, & Bresnick, 1993; Anderson & Sunder, 1995; Bolger & Wright, 1994; Johnson, 1995). As a result, expertise is context-dependent (Burgman, Carr, et al., 2011) and should be "unequally distributed" (including traditional and nontraditional experts) rather than merely determined by formal qualifications or professional membership (Evans, 2008). In response to the potential benefits of using expertise across the community (Carr, 2004; Hong & Page, 2004), Collins and Evans (2007) recommend the following prescriptions for identifying experts:

1. Identify core expertise requirements and the pool of potential experts, including lay expertise.
2. Create objective selection criteria and clear rules for engaging experts and stratify the pool of experts and select participants transparently based on the strata.
3. Evaluate the social and scientific context of the problem.
4. Identify potential conflicts of interest and motivational biases and control bias by "balancing" the composition of expert groups with respect to the issue at hand (especially if the pool of experts is small).
5. Test expertise relevant to the issues.
6. Provide opportunities for stakeholders to cross-examine all expert opinions.
7. Train experts and provide routine, systematic, relevant feedback on their performance.

Furthermore, Cooke and Goossens (2000) noted that experts should be willing to be identified publicly (but their exact judgments may be withheld except for competent peer review), provide their rationale supporting their judgments, and disclose any potential conflicts of interest. Cooke and Goossens recommended the following procedure:

1. Publish expert names and affiliations in the study.
2. Retain all information for competent peer review (post-evaluation), but not for unrestricted distribution.
3. Allow de-identified judgments to be available for unrestricted distribution.
4. Document and supply rationales for all judgments.
5. Provide each expert with feedback on his or her own performance.
6. Request expert permission for any published use beyond the above.

Though this procedure was devised specifically for studies that would guide public policy (where validation and transparency are important), similar measures could be employed for project management.

Once the pool of experts with the requisite expertise has been established, then an appropriate number of experts must be selected. The number of experts selected depends upon the nature of the decision context and the nature of the problem, including the degree of uncertainty expected. As a general rule of thumb, six to eight experts (and no fewer than four) (Clemen & Winkler, 1999; Hora, 2004) should be obtained, and at least some of the experts should come from outside of the organization conducting the elicitation. A pool of candidate experts (who possess the requisite expertise and have demonstrated interest and commitment to participate) may be reviewed by a committee, and a sufficient number of the best experts should be selected from that pool.

Experimental research has shown that expert performance is also impacted by the format of the elicitation process (Aloysius, Davis, Wilson, Taylor, & Kottwmann, 2006; Bottomley, Doyle, & Green, 2000; Fong et al., 2015). Therefore, from the planning stage, the form of the information elicited must be considered (i.e., whether it is generative or evaluative). Once the form of the expert judgment has been identified as either generative or evaluative, the requisite expertise must be identified and an adequate pool of experts must be selected to achieve that expertise. It was initially suggested that numerate experts would perform better on quantitative (or evaluative) exlicitation tasks, while literate experts would perform better on qualitative (or generative) elicitation tasks (Larichev & Brown, 2000). More recently, this was demonstrated using technically equivalent numerical and nonnumerical elicitation methods (Fasolo & Bana e Costa, 2014). Using this framework, Figure 5 provides a mapping of the various families of expert judgment elicitation methods to the two types of expertise.

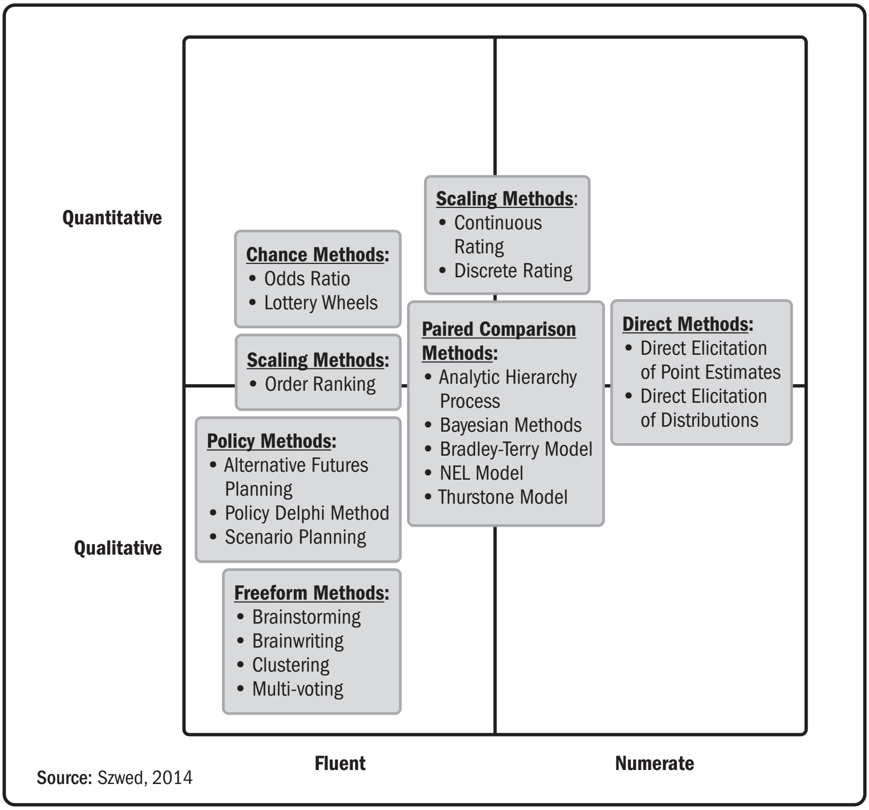


Figure 5 Taxonomy of expert judgment elicitation methods.

Numerate experts have facility with and possess the ability to discuss and describe quantities, probabilities, and numbers. Fluent (or literate) experts possess the ability to discuss and describe qualities using words and narrative. Some experts may be both numerate and literate. Additionally, in some contexts, experts may be more comfortable providing relative estimates. In such cases, paired comparison methods and scaling methods using order ranking would be appropriate. When experts are capable of providing absolute estimates, direct methods and scaling methods using discrete or continuous ratings would be appropriate. Decisions (or generative processes in the project management world) have been referred to as nodes for creativity (Kara, 2015), and fluency has been indicated as a means to invoke both system 1 and system 2 (Smerek, 2014). Fluency is often measured by examining vocabulary and testing the number of words an expert can spontaneously generate that begin with a specific letter or letters in a limited amount of time (e.g., Guilford, 1967; Guilford & Guilford, 1980; Spreen & Strauss, 1998). In cases where generative methods are deemed most appropriate, fluent (or literate) experts should be chosen. Likewise, in cases where evaluative methods are deemed most appropriate, numerate experts should be sought.

Because a significant portion of the expert judgment elicitation processes is evaluative in nature, it would be beneficial if there could be some means to assess experts' capabilities within system 2 (as shown in Figure 4). If the elicitation process called for evaluation that required system 2 cognitive skills and an expert unconsciously relied on system 1 to develop judgments (which we know are prone to a great many cognitive biases and heuristics), we would want some means to evaluate which system the expert invoked.

One of the main functions of system 2 is to monitor and control the thoughts and actions "suggested" by system 1, allowing some to be expressed directly in behavior and suppressing or modifying others. Because awareness about how judgment is processed is not readily apparent, Frederick (2005) developed the Cognitive Reflection Test to determine whether someone was actively employing system 2. The following questions come from this test:

1. A bat and ball cost US$1.10. The bat costs one dollar more than the ball. How much does the ball cost?
   a. 10 cents
   b. 5 cents
2. It takes 5 machines 5 minutes to makes 5 widgets. How long would it take 100 machines to make 100 widgets?
   a. 100 minutes
   b. 5 minutes
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take the patch to cover half of the lake?
   a. 24 days
   b. 47 days

In each of these questions, the correct answer is the second one; however, our intuitive system 1 response would have us favoring the first answer. Because we are overconfident in our intuitions, we often fail to check our work. Thousands of college students have been given this test. More than 80% (50% at more selective schools) gave the intuitive—incorrect—answers. Even though the answers are easily calculated, students who answered incorrectly simply did not check their work and relied on their intuition. Thus, it would be beneficial to use such a simple diagnostic test, in addition to subjective expertise, to determine which experts demonstrate the most control over their intuitions (and thus may be less susceptible to bias) (see, e.g., Tumonis, Šavelskis, & Žalytė, 2013). Campitelli and Labolitta (2010) found that cognitive reflection is related to the concept of actively open-minded thinking and that it interacts positively with knowledge and domain-specific heuristics and plays an important role in the adaptation of the expert to the decision environment. Cognitive reflection was also found to be a better predictor of performance on heuristics-and-biases tasks than cognitive ability, thinking dispositions, and executive functioning (Toplak, West, & Stanovich, 2011). Thus, it is expected that cognitive reflection will be a useful screening technique for experts. Additionally, by being better able to retrieve and use applicable numerical ideas, highly numerate experts have been shown to be less susceptible to biases (such as framing effects) and also more effective at providing numerical estimates (Lipkus, Samsa, & Rimer, 2001).

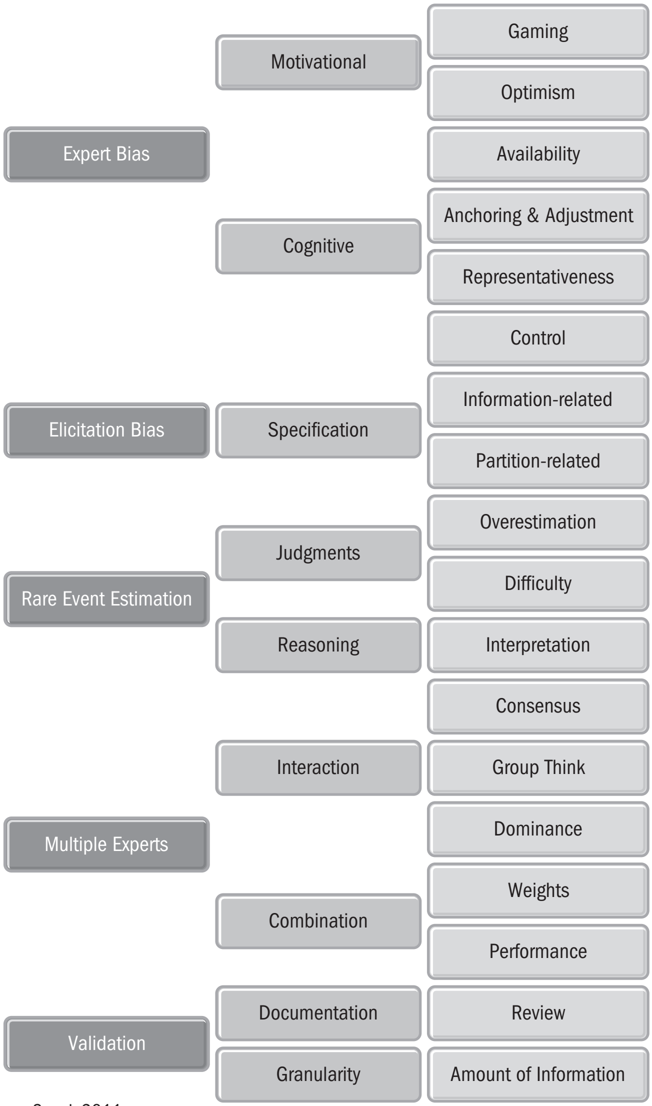
All of these considerations—requisite expertise, number of experts, form of expert judgment, and expert performance—will shape the selection of the pool of experts.

2.4.3 Training Experts

Once experts have been selected, they should be trained in an effort to ensure and improve the quality of their judgments. Expert judgment is influenced by many known issues and challenges (see Figure 6).

As a result of the known issues, there are several reasons for conducting pre-elicitation training of experts:

- To familiarize the experts with the problem under consideration and ensure that they share a similar baseline of information (e.g., basic domain knowledge or probabilistic and uncertainty training)
- To introduce the experts to the elicitation protocol, procedure, and process
- To introduce or reinforce the experts on uncertainty and probability encoding and provide them practice in formally articulating their judgments and rationale
- To provide awareness of the potential for cognitive biases that may influence their judgments



Source: Susel, 2011

Figure 6 Taxonomy of expert elicitation issues.

Fischhoff (1982) proposed a framework for elicitation enhancement strategies. These were simplified by O'Hagan et al. (2006) into the following:

- Fix the task.
- Fix the expert.
- Match the expert to the task.

We will examine each of these in reverse order. Matching the expert to the task at hand was covered in detail in the previous section. In short, qualitative/quantitative information should be elicited using generative/evaluative methods to gather judgment from fluent/numerate experts. Next, we will examine some of the efforts that work to "fix" or train the expert. Following that, there will be a considerable look at "fixing" the task—typically by creating methods for minimizing expert bias.

Before the elicitation session begins, it is important to explain to the experts why their judgments are required. Clemen and Reilly (2001) note that it is important to establish rapport with the experts and to engender enthusiasm for the project. Walker, Evans, and MacIntosh (2001) suggest that training of experts should involve:

- information about the judgments (e.g., probability distributions);
- information about the most common cognitive biases and heuristics, including advice on how to overcome them; and
- practice elicitations (particularly examples where the true value is known).

In other words, if it is possible, you would like the experts to share a common understanding of exactly what information is being elicited. Although experts will approach the elicitation with a variety of differing perspectives based upon their diversity of training and experience, it is paramount that they all address the same problem as posed by the elicitation. This can be accomplished through pre-elicitation training. Also, it is important to allow the experts to gain experience with the elicitation protocol (i.e., the questionnaire, survey, interview, etc.) in advance of the actual elicitation. This way, when it comes time to provide their judgments, they will have the best chance of being consistently supplied.

Pre-elicitation training may also include tuning expert numeracy—for example, many experts are not familiar with describing their degrees of belief and uncertainty in terms of quantiles (e.g., 5%, 50%, 95%). Allowing all experts to participate in a group training session allows each the benefit of hearing the others' questions (and your responses) and ensures that all have a common understanding of what will be asked of them.

In terms of expert judgment elicitation, there are a number of common mechanisms that have been used for debiasing (Jørgensen, Halkjelsvik, & Kitchenham, 2012; Simola, Mengolini, Bolado-Lavin, & Gandossi, 2005) experts (typically in attempts to reduce overconfidence):

- Expert training
- Feedback
- Incentive schemes, such as scoring rules

Though these efforts have been met with mixed results in the past (Alpert & Raiffa, 1982; Arkes, Christianson, Lai, & Blumer, 1987; Hogarth, 1975; Koriat, Lichtenstein, & Fischoff, 1980; Lichtenstein et al., 1981), more recent efforts (which will be described next) have demonstrated results in helping to debias experts.

Considerable attention has been devoted to the challenge brought about by cognitive biases and heuristics. For in-depth coverage of this specific set of issues, please refer to any of the many comprehensive books on the subject (e.g., Gilovich, Griffin, & Kahneman, 2002; Kahneman et al., 1982; Kahneman & Tversky, 2000; Tversky & Kahneman, 1974). Some of the most common biases and heuristics are described in Table 8.

Table 8 Cognitive biases and heuristics.

Bias or Heuristic	Description	Solution
Anchoring	Anchoring-and-adjustment involves starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, "**adjustments are typically insufficient**" (Tversky & Kahneman, 1974, p. 1128).	The most common demonstration of anchoring would be when a three-point estimate is asked of the expert: the most likely value or 50%, the practical minimum or 5%, and the practical maximum or 95%. Rather than asking for the most likely first, ask for the range values, the 5% and 95% (such that there will be a 90% chance the true value falls into the range), and then the most likely, the 50%. This will help avoid insufficient adjustment from the most likely estimate.
Representativeness	Representativeness refers to making an uncertainty judgment on the basis of "the degree to which it (i) is similar in essential properties to its parent population and (ii) reflects the salient features of the process by which it is generated" (Kahneman & Tversky, 1972, p. 431). Supporting evidence has come from reports that people ignore base rates, neglect sample size, overlook regression toward the mean, and misestimate conjunctive probabilities (Kahneman & Tversky, 2000; Tversky & Kahneman, 1974).	The representativeness bias can become an issue when extrapolating data or judgments from known populations of different sizes. This is particularly prevalent when making comparisons of likelihood. As a result, it is important to remind experts to continually think about base rates, sample size, and regression to the mean.
Availability	Availability is used to estimate "frequency or probability by the ease with which instances or associations come to mind" (Tversky & Kahneman, 1974, p. 208). In contrast to representativeness, which involves assessments of similarity or connotative distance, availability reflects assessments of associative distance (Tversky & Kahneman, 1974). Availability has been reported to be influenced by imaginability, familiarity, and vividness, and has been supported by evidence of stereotypic and scenario thinking (Tversky & Kahneman, 1974).	Availability results from an ease of recall. One way to mitigate effects of the availability heuristic would be to require and allow experts to review and consider the full spectrum of reports and studies immediately prior to the elicitation. This will enable them to have all of the information available for their judgments rather than just those that are more recent in memory.
Framing	Drawing different conclusions from the same information based on how that information is presented (Tversky & Kahneman, 1981). For example, suppose a scenario is presented such that an outbreak of an unusual disease is expected to kill 600 people. Two alternative programs are suggested. When the difference between the programs was framed showing program A saved 200 people and program B had P(600 saved) = 1/3 and P(0 saved) = 2/3, 72% of respondents opted for program A. However, when the difference between the programs was framed showing program C where 400 people died and program D had P(0 die) = 1/3 and P(600 die) = 2/3, 78% of respondents opted for program D. Even though programs A and C are identical (as are programs B and D), the results were different based on framing.	One way to mitigate the effects of framing is to carefully use neutral wording. Another possible way is to provide equivalent wordings to demonstrate potential framing issues.
Overconfidence	Overconfidence results when an expert's subjective confidence in their own judgments exceeds (or is reliably greater than) their objective accuracy. This overconfidence can be observed in subjective statements of confidence or when the range of the 5% and 95% estimates of a three-point estimate are insufficiently broadly ranged, and thus the standard deviation or variance of their distribution is too small. Overconfidence may be observed when it is possible to evaluate expert performance using known seed variables.	Overconfidence may be held partially in check by demonstrating propensity for overconfidence during training. For example, experts will be asked to provide three-point estimates (5%, 50%, and 95%) for five known encyclopedic quantities (e.g. length of Mississippi River, population of Washington, DC). Typically, less than 5% of experts will answer all five questions such that the true value falls within their range of confidence (between 5% and 95%). The majority of experts will correctly capture within their range the true value of two or fewer quantities. This training causes experts to more accurately express their uncertainty and better calibrate their confidence.

Looking only at a single bias, in 2011, a metasearch of more than 100 electronic bibliographic databases identified 2,092 articles with “overconfidence” in the title for an eight-year period. Of those, 26 (some with duplicates) clustered on the keyword “interval estimates.” A survey of the most recent research on overconfidence and interval estimation will now be summarized using the taxonomy provided in Figure 7. Articles generally fell within two categories: ones that dealt with improving elicitation via various methods and others that used feedback to improve elicitation. The majority of the “method” articles were focused on interval estimation, but there were two additional methods (i.e., SPIES and MOLE). The interval estimation methods were then broken down into whether or not the interval was specified in advance (thus, pre-designated) or whether the experts assigned their own intervals. Finally, the methods were further broken down into how many estimates or “points” were required for each judgment. The “feedback” articles were primarily sorted by the source of the feedback (i.e., self when experts were provided feedback on their own estimates, others when experts were provided feedback on other experts, estimates, and actual when feedback provided the eventual actual value of the estimate).

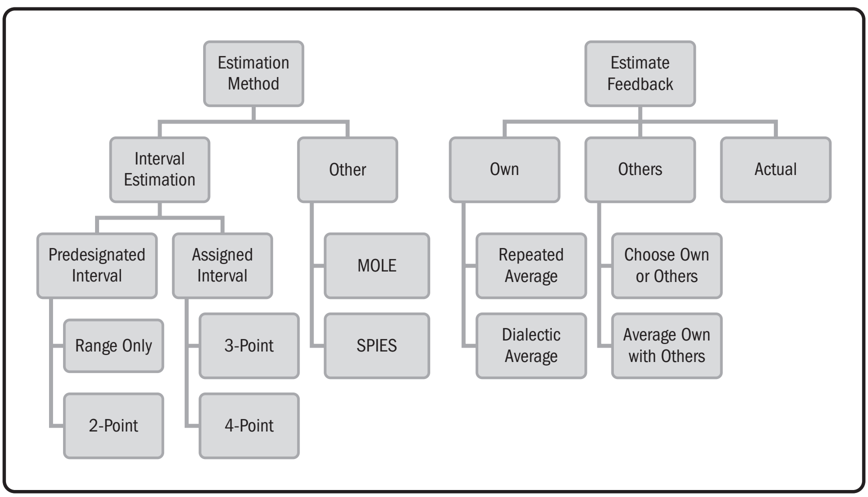


Figure 7 Taxonomy of some expert elicitation research.

Interval estimation is a widely used means of eliciting expert judgments. Typically, experts are asked to provide values for a predesignated interval. For example, an expert would be asked to provide the 5th and 95th percentile estimates (such that the true value would theoretically occur within that interval 90% of the time, in the long run). Soll and Klayman (2004) examined the effects of a range only (so-called one-point, two-point [where experts are asked above and below percentile values], and three-point estimates [which include an estimate of the most likely value in addition to the estimates assigned to the upper and lower bounds of the predesignated interval]). They demonstrated that the three-point estimate produced an average overconfidence of 14%, in comparison to average overconfidence of 23% and 41% for two-point and one-point estimates, respectively.

An alternative to this approach is to ask the experts to estimate the practical minimum and the practical maximum, and then ask them to assign an interval based upon their confidence (e.g., 50%, 82%, and 90%). Winman, Hansson, and Juslin (2004) have shown that allowing experts to assign the interval (what they called interval evaluation) yielded less overconfidence than having experts produce a predetermined interval. Average overconfidence dropped from 32% for the estimates of predesignated intervals to 14% for expert-assigned intervals.

Teigen and Jørgensen (2005) conducted similar experiments and demonstrated similar improvements in overconfidence reduction. Estimates of the predesignated 90% interval had an average overconfidence of 67%, estimates of the predesignated 50% interval had an average overconfidence of 27%, and estimates for the expert-assigned intervals had an average overconfidence of 15%. Speirs-Bridge et al. (2010) extended this investigation to include a four-point estimate. The four-point estimate asked the experts to identify upper and lower bounds, in addition to the most likely value, and then asked the experts to assign an interval based upon their confidence. The study employed authentic experts estimating real information (as opposed to students estimating artificial values) and yielded improvements in overconfidence; the four-point method had an average overconfidence of 12%, compared with 28% for the three-point method. Thus, one conclusion to be drawn is that when experts are asked to provide additional information about their estimates (from one to two to three and finally to four points), overconfidence is reduced.

The more-or-less-elicitation (MOLE) method has demonstrated improved accuracy and precision for elicited ranges (Welsh et al., 2008, 2009), but requires repeated relative judgments that are not often possible in the context of extensive expert elicitations. Similarly, subjective probability interval estimates (SPIES) require experts to assign probabilities across the full range of possible values; this method yielded 3% overconfidence, as compared to 16% overconfidence using a three-point estimate (Haran et al., 2010). This may not be feasible either; first, the range of possible values may not be known, and second, requiring multiple elicitations for each quantity of interest to obtain a full distribution will likely exceed the cognitive capacity of experts and result in fatigue during extensive elicitations. Therefore, although the MOLE and SPIES methods show promise, the focus here will return to interval estimation.

Alternatively, aside from the method chosen, how the problem is decomposed is also important. There is emerging evidence that “unpacking the future” by decomposing the distal future into more proximal futures improves calibration and reduces overconfidence (Jain, Mukherjee, Bearden, & Gaba, 2013).

Feedback regarding actual performance on elicitations will improve overconfidence. When experts were provided feedback on how well their estimated intervals compared to the true values, overconfidence was reduced from 16% to 2% after the first session and to -4% after multiple sessions (Bolger & Önkal-Atay, 2004).

Furthermore, the average of quantitative estimates of a group of individuals is consistently more accurate than the typical single estimate because both random and systematic errors tend to cancel (Vul & Pashler, 2008), a phenomenon that has become known as the wisdom of the crowd. Additionally, a similar effect can be created when one individual makes repeated estimates. Averaging a first estimate with a second, dialectic (i.e., antithetical) estimate simulates an averaging of errors (Herzog & Hertwig, 2009); even though averaging with another expert increased accuracy by 7%, the dialectic averaging increased accuracy by 4% (beyond that from mere reliability gains). Although it has been shown that diverse groups make better decisions than individuals (or homogeneous expert groups) (Hong & Page, 2004), the internal averaging effect was confirmed in another study of the "wisdom of crowds in one mind" and calculated the optimal number of times to elicit from each individual expert (Rauhut & Lorenz, 2010). When given the choice of whether or not to average with other experts, those who chose their own judgments over others' frequently exhibited 24% overconfidence, those who occasionally chose their own over others' exhibited 17% overconfidence, and those who regularly combined their judgments with others' exhibited 13% overconfidence (Soll & Larrick, 2009). Therefore, feedback and averaging generally help reduce overconfidence.

Even with this quick review of a small slice of the expert judgment literature, it should be readily apparent that there is a considerable amount of experiments, study, literature, and theory on expert judgment that should ultimately be applied to the practice of project management in a more systematic manner.

Additionally, because bias may have the greatest impact on judgment, attention can be focused on debiasing. Ways to debias judgment include "modifying" either the person or the environment (Soll, Milkman, & Payne, 2014), including teaching cognitive strategies, providing nudges to induce reflective thinking (Jain et al., 2013), and so on. Training could also be as simple as providing the experts with exposure to the types of questions that will prime their thinking.

Appropriate expert training may include a mix of orientation, practice, debiasing, feedback, and providing appropriate incentives.

2.4.4 Eliciting Judgments

As seen previously, there are many elicitation methods for gathering expert judgments. Also, it has been observed that the method must be matched to the purpose or problem.

> Due to the subjective nature of elicitation it is important to provide a transparent account of how values are elicited and what information was available to experts to aid in their estimation of various quantities. (Roelofs & Roelofs, 2013, p. 1651)

For example, if the elicitation task involves creating a list of ideas, scenarios, risk, and so forth, then a *generative* method would be the best choice. As noted in the state-of-the-practice survey, brainstorming was the most frequently used expert judgment tool/technique. It is easy to implement, familiar, and widely used. Osborn (1957) noted that people can generate twice as many ideas when working in groups compared to working alone by adhering to the following simple rules: More ideas are better, wilder ideas are better, improve and combine ideas to create more, and refrain from criticism. These rules remove the inhibitions of criticism. However, researchers have demonstrated that nominal groups (where participants work independently before combining their ideas) outperform brainstorming groups by a factor of two under similar conditions (e.g., Taylor, Berry, & Block, 1958). Some of the theories of the productivity loss from brainstorming include *production blocking* (from monochannel communication) (Lamn & Trommsdorff, 1973), *evaluation apprehension* (or fear of criticsm from others) (Collaros & Anderson, 1969), and *free riding* (or social loafing). The effects of these barriers to productivity may be moderated through allowing independent work prior to group brainstorming sessions, clearly describing instructions and rules against evaluation, limiting group size, and ensuring incentives and assessment evaluation (Diehl & Stroebe, 1987). Other more recent techniques to improve brainstorming include cognitive priming (Dennis, Minas, & Bhagwatwar, 2013), avoiding categorization or clustering a priori (Deuja, Kohn, Paulus, & Korde, 2014), and pacing/awareness to avoid cognitive fixation (Kohn & Smith, 2011).

Despite its prevalence and persistence (Gobble, 2014) and the fact that it may well serve other organizational purposes (Furnham, 2000), brainstorming will continue to be an option and should be used with due caution and in accordance with the intent of the original creators and the researchers who have improved the productivity of brainstorming.

There are many variants of the traditional (now almost six decades old) technique of brainstorming. Increasingly, brainstorming involves virtual groups (e.g., Alahuhta, Nordbäck, Sivunen, & Surakka, 2014; Dzindolet, Paulus, & Glazer, 2012) and crowd sourcing (Poetz & Schreier, 2012), as technology improves opportunities for involvement from a diversity of experts and users. Brainwriting (or brain sketching) is another method of generative elicitation, where group members begin by silently sketching their ideas and annotations on large sheets of paper that are then shared among group members for another round of brainwriting (VanGundy, 1988). Six-five-three brainwriting is where six participants each write three ideas on a sheet of paper that is circulated and the process is repeated five times (e.g., Otto & Wood, 2001; Shah, 1998; Shah, Kulkarni, & Vargas-Hernández, 2000; Shah, Smith, & Vargas-Hernández, 2003). Collaborative sketching (or C-Sketch) (Shah, Vargas-Hernández, Summers, & Kulkarni, 2001) is a variant of 6-5-3 brainwriting, where participants draw diagrams instead of using words, and the misinterpretation of ambiguous drawings may lead to new ideas. One study found rotational brainwriting techniques to outperform nominal group techniques in terms of both quantity and quality (Linsey & Becker, 2011).

Another common method, the nominal group technique (Delbecq & Van de Ven, 1975), has proven to be equally or more effective than brainstorming. This technique involves having group members work in silence without talking (i.e., they are a group in name only) before ideas are shared and expanded.

As can be imagined, there are hundreds of idea generation techniques (e.g., Adams, 1986; Higgins, 1994). There is no single best method (e.g., nominal groups outperform brainstorming, brainwriting outperforms nominal groups, different methods outperform others in differing contexts, etc.), and the selection of methods will be dictated by the nature of the problem and the skill and experience of the person(s) conducting the elicitation.

When it comes to evaluative expert elicitation, there are even more methods to choose from (see Figure 5 and Table 6). Additionally, several prominent methods are described in this chapter. It would be far too cumbersome to attempt to explain all of the various methods here. Instead, attention will be turned toward the elicitation process itself (rather than the method of elicitation) to identify some of the key elements of an elicitation according to Meyer and Booker (2001):

- Will the elicitation be individual or interactive (i.e., the situation or setting)?
- What form of communication will be used (e.g., face-to-face, virtual, etc.)?
- Which technique will be selected (see Tables 5 and 6 for examples of the various generative and evaluative methods)?
- What will be the form of the response mode (e.g., estimate, rating, ranking, open, etc.)?
- Will experts be provided with feedback?

There are many factors to consider when designing an expert judgment elicitation (e.g., how best to survey experts [Baker, Bosetti, Jenni, & Ricci, 2014]) and, in order to achieve the best possible expert judgment, much planning and attention must be paid to the execution.

2.4.5 Analyzing and Aggregating Judgments

Once the judgments have been elicited, they will need to be evaluated and, if deemed necessary, combined or aggregated.

The nature of how the judgments are analyzed and aggregated will be dependent upon the form of the information or data sought. There are two basic forms of aggregation methods: behavioral and mathematical. In general, generative (qualitative) judgments are most often combined using behavioral aggregation methods, and evaluative (quantitative) judgments are combined using mathematical aggregation methods.

There are several comprehensive reviews of aggregation methods in the literature (e.g., Clemen, 1989; Clemen & Winkler, 1999; French, 1985, 2011; Genest & Zidek, 1986), including several that have annotated bibliographies. Rather than replicate those contributions here, only the high points will be summarized.

Behavioral methods require the experts to interact in order to generate some agreement. These methods generate consensus aggregation as a by-product of the expert judgment elicitation process, rather than as a consequence of some manipulation after the elicitation (as is the case with mathematical aggregation). The following are some behavioral expert judgment aggregation methods:

- *Group Assignment:* This method has the experts work together to develop a group assignment of the probability distribution or quantity of interest.
- *Consensus Direct Estimation:* Here, too, the group of experts identifies a quantity of interest and then comes to a consensus through interaction.
- *Delphi Method* (Dalkey & Helmer, 1963): This well-known iterative, asynchronous process is typically performed by experts independently. The anonymous results are then shared and other experts are allowed to comment and update their estimates. Rounds continue until there is sufficient consensus. Its advantages include anonymity, the opportunity to gain new information or defend outlier position, and self-rating. There are also many disadvantages, as enumerated by Sackman (1975), including the fact that the process

is time-consuming, does not adhere to psychometric rules, results in unequal treatment of experts, offers no means for dealing with lack of consensus, requires no explanation as to why experts prematurely exit surveys, and convergence may be the result of boredom. Gustafson, Shulka, Delbecq, and Walster (1973) found that the technique produced worse results than the nominal group technique and simple averaging.

- *Nominal Group Technique* (Delbecq & Van de Ven, 1975): Experts make judgments first independently and then come together to form consensus. This technique allows synergy among experts, but there is potential for bias (as in any of the behavioral methods).
- *Analytic Hierarchy Process* (Saaty, 1980): Experts individually rank alternatives using relative scales and the rankings are then combined. The advantages of this process are that it allows hierarchical design, is easy, is structured such that comparisons are reciprocal, and provides means for diagnosing experts through consistency measures. The disadvantages have been identified by Dyer (1990) and include the potential for rank reversal and independence in weights between hierarchies.
- *Kaplan Method* (Kaplan, 1990): This method requires the facilitation of experts in discussing and developing a consensus body of evidence. Using that consensus body of evidence, a distribution is proposed and then argued among the experts based upon the shared evidence until consensus is obtained.

With all interactive groups, there is potential for problems, such as groupthink (Jannis, 1982; Jannis & Mann, 1977), polarization (Plous, 1993), and expert dominance. Despite these potential problems (which can be addressed through the elicitation protocol), group performance is typically better than that of the average group member, but not as good as that of the best group member, according to one study of 50 years of research on decision making (Hill, 1982).

Mathematical methods use analytical processes or mathematical models to combine individual expert judgments into a combined judgment. The following are some mathematical expert judgment aggregation methods (which are typically conducted after judgments have been elicited):

- *Weighted Arithmetic Mean* (also known as Linear Opinion Pool [Stone, 1961]): This appealing approach averages expert judgments (e.g., probability) using the same or performance-based weights. It is a simple average of the judgments. Some advantages of this method include ease of calculation, ease of understanding, maintaining unanimity, the fact that weights can represent expert quality, and satisfying marginalization. However, the determination of weights may be subjective (Genest & McConway, 1990).
- *Weighted Geometric Mean* (also known as Logarithmic Opinion Pool): This method uses a multiplicative average and a normalizing constant. It is a weighted average. Some advantages include that it can be easily updated with new information and weights can represent expert quality. Again, determination of weights is subjective.
- *Mendell-Sheridan Model* (1989): This Bayesian approach creates joint expert quantile estimates. In contrast to frequentist approaches to determining probabilities based upon the frequency of occurrence of an event, Bayesian approaches allow a degree of belief to be incorporated. Some of the advantages include that it has default egalitarian priors (i.e., equally weighted prior beliefs), experts are not restricted to a class of distribution, it updates

accounts for correlation, and it has been experimentally tested. It is, however, computationally complex, sensitive to units, and dependent upon seed variables to "warm up."

- *Morris Model* (1977): This Bayesian method provides a composite probability assignment for the quantity of interest incorporating the decision makers' prior understanding of the situation (offered in the form of a distribution). Some advantages include the fact that conflicting expert assessments can be accommodated, invariance to scale and shift, precision and accuracy are dealt with separately through decomposition, and calibration is inherent to the model. Disadvantages include its restriction to normality assumptions of expert priors and the fact that it does not address the issue of expert dependence.
- *Additive (and Multiplicative) Error Models* (Mosleh & Apostalakis, 1982): Using vectors for quantile estimates, a combined distribution is developed that provides expert performance and correlation data. It is relatively simple and the errors are normally distributed. However, there is a heavy burden on the decision maker to supply the prior, bias, and accuracy for each expert. Also, it does not generalize to all classes of distributions.
- *Paired Comparison Model* (Pulkkinen, 1993): This Bayesian model creates a composite posterior mean and variance based upon expert-paired comparison information. Advantages include the fact that the comparisons are intuitively accessible, the likelihood is derived from comparison responses, and it is relatively flexible. However, it has weak dependence among experts and cannot be solved in closed form (requiring simulation solution).
- *Information Theoretic Model* (Kullback, 1959): This method identifies an aggregate probability distribution

with the least cross-entropy. Some advantages are that it is a normative model, it retains distributional family (often a simple combination of parameters), and it provides alternative objective criteria to fit application. However, experts may require weighting.

Many of the mathematical aggregation methods use performance-based weights (or scoring rules) as a means of combining expert judgments, such that the judgments of more accurate experts (i.e., those with superior performance) are given higher weighting in the aggregation (and vice versa). However, evaluating expert performance is extremely difficult, and the quality of the "experiential insight" (Crawford-Brown, 2001) of each particular expert must be evaluated as objectively as possible. This is difficult: The accuracy of an expert's judgment about an unknown quantity of interest is not typically known at the time of the aggregation because the project has not yet taken place. However, there are means to evaluate expert accuracy using "seed" variables. The actual values of the seed variables are known to the analysts administering the elicitation (e.g., in cases where historic data may be available or in cases where additional reports were unavailable to the experts). The seed variables are introduced into the elicitation protocol, and experts (who do not explicitly know the true value of the seed variables) estimate those values along with those of the quantities of interest. Expert performance is then evaluated by examining the experts' performance on the set of seed variables. Measuring expert performance is important because in addition to being a means of combining expert judgments, it can also serve to enhance the credibility of a study or plan. (See Cooke [1999] for a procedure that "calibrates" experts using seed variables.) Despite a strong case to be made for such an aggregation method, there is some evidence that Cooke's (1999) classical method performs no better than equal weighting and may suffer from sample bias (Clemen, 2008). There are also other metrics of expert performance using various forms of measuring accuracy, bias, and calibration, in addition to conventional scoring rules (e.g., Cooke, 2015; DeGroot & Fienberg, 1982; Murphy, 1972a, 1972b; Matheson & Winkler, 1976; Yates, 1994a, 1994b).

Also, there may be some instances when you would not combine the judgments, such as when there may be considerable difference in opinion and it is important to retain that distribution of judgment (Keith, 1996). In summary, though there is an enormous amount of research and literature devoted to the topic of the aggregation of expert judgments, behavioral methods are best suited for generative judgments and mathematical methods are best for evaluative judgments. Further, although there are many different and sophisticated methods for mathematical aggregation, simple averaging often outperforms other methods (Clemen & Winkler, 1999).

2.5 Findings and Implications

This review of the literature has uncovered several significant findings regarding how the state of the art/science can inform the practice of expert judgment in project management:

- There are a great many elicitation protocols. The generic seven-step protocol is a compilation of some of the most prominent and adheres to the five phases of project management.
- Expert judgment in project management conforms to two basic forms: generative and evaluative.
- Generative information may best be obtained from fluent (or literate) experts. Evaluative information may best be obtained from numerate experts.
- Expert selection is important and expertise should be matched to the information needed.
- There are some expert judgment elicitation methods that are better suited for generative tasks. Other expert judgment elicitation methods are better suited for evaluative tasks.
- Expert judgment can be improved through pre-elicitation training, debiasing, and feedback.

- Expert judgments can be aggregated using mathematical (e.g., arithmetic average) or behavioral (e.g., consensus) means. Often, the best way to combine evaluative judgment (e.g., interval estimates) is by simple averaging.

Given the breadth and depth of the state of the art/science (i.e., in disciplines other than project management), there is considerable opportunity to improve the practice in project management and advance how expert judgment is used as a tool/technique.

CHAPTER 3

State of the Practice

To better understand how expert judgment is employed in the project management context, a global study of project management professionals was conducted.

3.1 Method

This phase of the research examined the state of the practice of expert judgment in project management. Figure 8 illustrates the model being explored by this study.

A descriptive survey was selected as an effective means to gather information that is not easily observed (Buckingham & Saunders, 2004). The survey was developed in 2013 and underwent institutional board review and then blind peer review according to the sponsoring organization's grant process. (See Appendix A for a complete listing of the instrument.) The survey was designed to do the following:

1. Determine current expert judgment practices in the project management context.
2. Compare expert judgment practices across different industries and regions.
3. Elicit best or effective expert judgment practices.

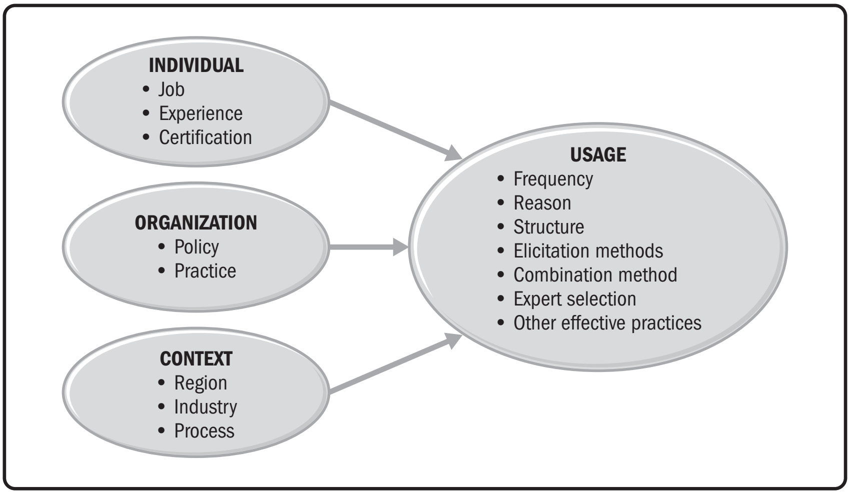


Figure 8 Research framework.

In order to achieve those objectives, two categories of questions were developed. The first group of questions gathered demographic information about the project management professionals (e.g., job function, experience, certification) and the context within which they work (e.g., industry, location, policy, and process). The second group of questions gathered expert judgment practice information (e.g., usage, process, tools).

3.2 Data and Sample

An online survey was administered during the second half of 2014. It was posted to the Project Management Institute's website during the first four months. The Project Management Institute is the largest global professional society for project and program managers with more than 450,000 members; it has 273 chartered chapters in 105 countries (Project Management Institute, 2015). Because there was insufficient response resulting from the passive deployment and because survey participation is declining in general (Kennedy & Vargus, 2001), in November and December, the survey was sent electronically to each of the chartered chapters requesting that the survey be shared with their local membership. This direct appeal, using a modified Dillman

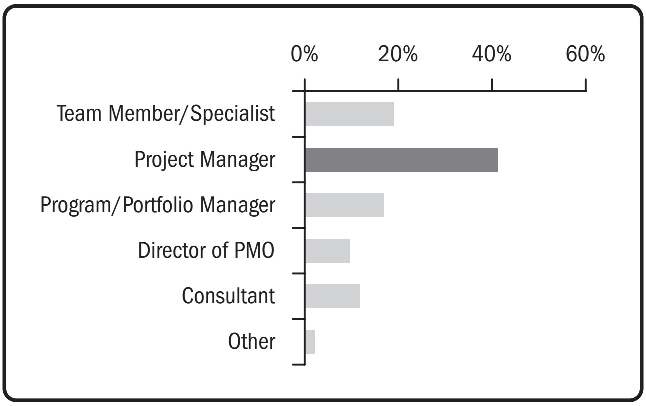


Figure 9 Respondent demographics—Primary job function.

technique (2011), yielded 449 responses to the survey, of which 382 were complete (i.e., a completion rate of 85.1%). Figures 9 through 12 provide a summary of the demographic data. The most frequently occurring responses (i.e., the modal response) have the darkest shading.

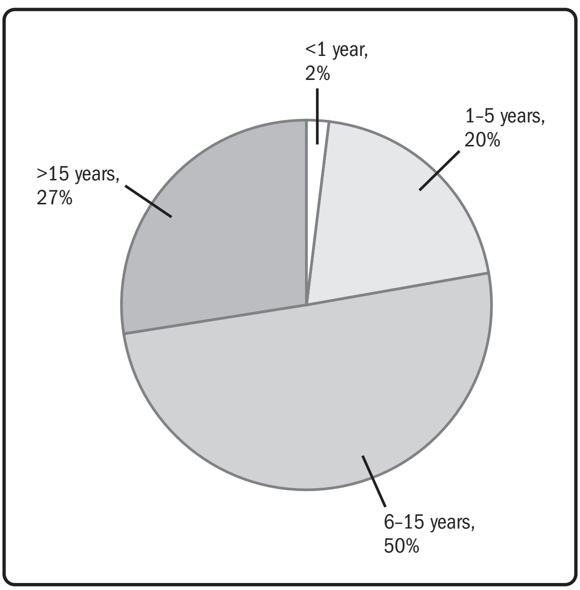


Figure 10 Respondent demographics—Project management experience.

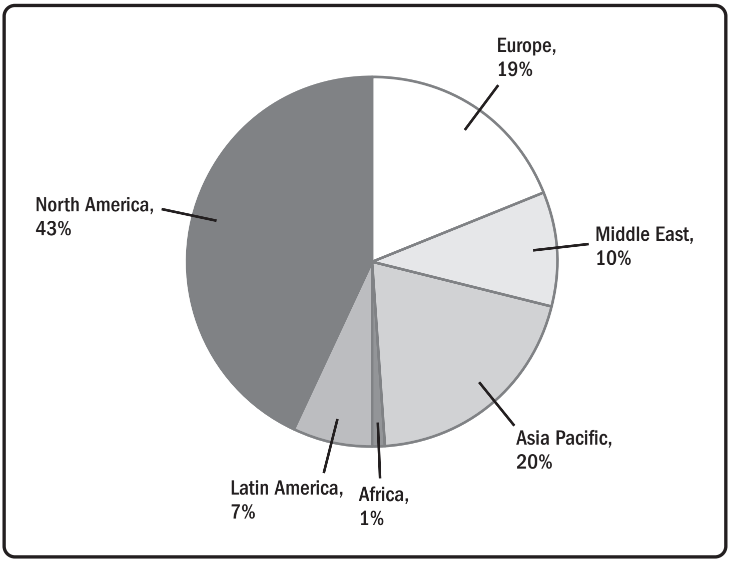


Figure 11 Respondent demographics—Region.

This sample is approximately representative of the population for the professional society membership (K. Dunn, personal communication, January 15, 2015). Table 9 describes the demographic breakdown of the sample as compared to the population from which the survey was drawn.

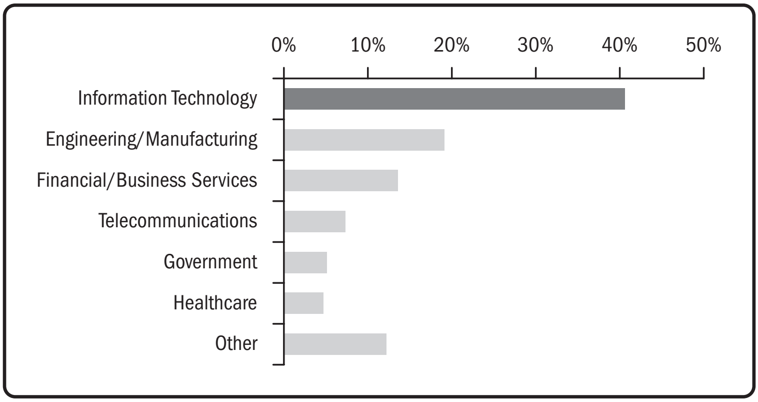


Figure 12 Respondent demographics—Industry.

Table 9 Demographics of sample compared to population.

	Sample (n = 382 respondents)	Population (N = 451,188 members)	Difference Between Sample & Population
Region			
• North America	42.2%	60.8%	(18.6%)
• Asia Pacific	18.7%	17.2%	1.5%
• Europe	17.2%	10.5%	6.7%
• Latin America & Caribbean	17.2%	7.1%	10.1%
• Middle East	3.7%	3.1%	0.6%
• Africa	1.0%	1.3%	(0.3%)
Experience			
• 6–15 years	51.2%	43.8%	7.4%
• >15 years	27.1%	30.7%	(3.6%)
• 1–5 years	20.3%	17.6%	2.7%
• <1 year	2.4%	7.9%	(5.4%)
Industry			
• Information Technology	40.5%	31.6%	8.9%
• Engineering & Manufacturing	18.6%	23.8%	(5.2%)
• Business & Financial Services	13.6%	13.5%	0.1%
• Telecommunications	6.5%	6.7%	(0.2%)
• Healthcare	4.7%	6.2%	(1.5%)
• Government	4.4%	4.0%	0.4%
• Other	11.7%	14.2%	(2.5%)

Note: Figures in parentheses represent negative differences between the sample and the population.

In general, the sample did not suffer coverage sampling error (Couper, 2000). The sample was slightly overrepresented in respondents from Europe and Latin America (and significantly underrepresented in North America—an increasingly common result [Cook, Heath, & Thompson, 2000]). The distribution of experience level of the respondents was representative of the population. Adjustments were made for the categories of experience to facilitate comparison. Additionally, the sample was moderately overrepresented in respondents from the information technology industry. Overall, the order of relative representation was consistent between the sample and the population, which illustrates an adequate sample (Groves et al., 2011).

3.3 Analysis and Results

Because this phase of the research set out to determine the state of the practice, the descriptive summary results were as important

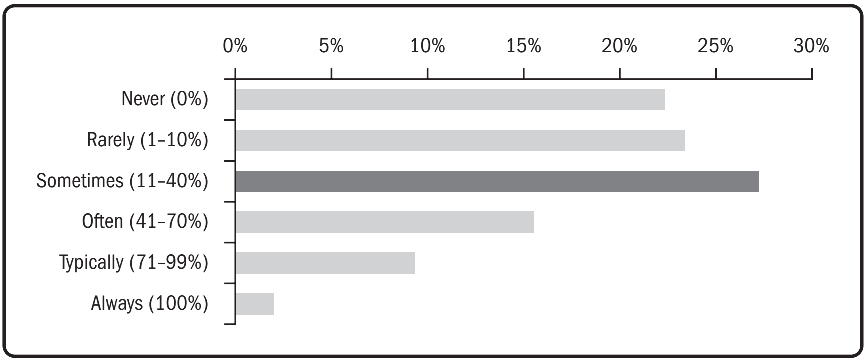


Figure 13 Portion of time structured expert judgment process is used.

as the analytical results. First, it was observed that a vast majority of respondents (95.1%) indicated using expert judgment for the projects they manage.

Since one of the underlying motivations of this entire research project was to determine if the practice of expert judgment in project management is ad hoc and ill-defined (as suggested by the limited definition), one of the most important questions in the survey was "How often do you use a predefined structured process for eliciting expert judgment?" Figure 13 provides a summary of responses to that question.

Respondents were provided a scale of frequency (on the lefthand side of Figure 13) with which they indicated how often they used a structured process for the elicitation of expert judgment. The majority of respondents (almost three quarters) used a structured process infrequently or not at all. If these responses were combined, we would see that expert judgment is only elicited 25.5% of the time using a predefined structured process. Therefore, expert judgment is not usually elicited using a predefined structured process. Overall, just one in 10 project management professionals uses written guidance for eliciting expert judgment.

Figure 14 describes whether or not the organizations within which the respondents worked offered written guidance on expert judgment elicitation.

Almost four out of every five respondents worked in organizations that did not have any written guidance on expert judgment. Of those who responded that they did work in an organization that had written guidance, only about 60% indicated that the guidance was typically followed. The other roughly 40% indicated that the guidance was rarely followed, effectively indicating that 87.8% of respondents had not used written guidance about eliciting expert judgment.

Several of the hypotheses suggested relationships between demographic variables and practice variables. Analysis of variation (ANOVA) was performed, and Table 10 provides the correlation matrix. All of these values were tested at the 95% significance level.

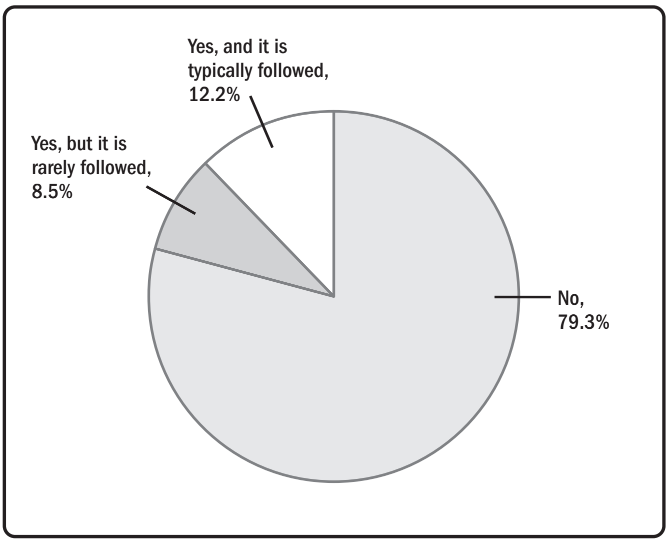


Figure 14 Portion of organizations with written expert judgment guidance.

Table 10 Correlation matrix.

	JOB	EXPERIENCE	CERT	INDUSTRY	REGION	USE	POLICY	STRUCTURE	COMBINE
JOB	#####								
EXPERIENCE	30.1%	#####							
CERT	3.9%	17.2%	#####						
INDUSTRY	–6.2%	–7.3%	9.8%	#####					
REGION	–1.6%	15.6%	4.1%	–5.4%	#####				
USE	0.5%	13.1%	16.7%	0.8%	5.6%	#####			
POLICY	5.5%	0.0%	–8.1%	–8.5%	–4.7%	10.1%	#####		
STRUCTURE	7.8%	2.2%	–7.8%	0.3%	–3.3%	10.2%	50.9%	#####	
COMBINE	–9.1%	–3.7%	0.6%	1.0%	–7.8%	14.7%	16.5%	16.9%	#####

There was not a significant relationship between project management experience and frequency with which a predefined structured elicitation process was used. Therefore, there was insufficient evidence to support the idea that the usage of a predefined process was independent of and not correlated to project management experience. Note also that there was no correlation between the region where the respondent project manager worked and whether a predefined structured process for eliciting expert judgment was used. Therefore, there was insufficient evidence to support the notion that the usage of a predefined process was independent of and not correlated to the region in which the project manager worked.

Observing the strongest correlations, the greatest is the 50.9% association between whether or not an organization has written guidance on expert elicitation and how often a project manager uses a predefined structured process for eliciting expert judgment. This was the strongest correlation and marginally upheld the notion that the use of a predefined structured process increases when an organization has written guidance for conducting expert judgment elicitation. The next strongest was between project management experience and job function. Although not a part of this study, this relationship could be expected because as project managers gain experience, they likely move from serving on a team to leading a team to leading several teams to leading an organization.

Figure 15 provides an overview of the most frequently used expert judgment elicitation methods; brainstorming and direct estimation were the two most frequently used.

The Project Management Institute publishes a set of guidelines that includes standard definitions, terminology, processes, and procedures for conducting project management. The most recent edition of this 589-page foundational standard, known as *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)* – Fifth Edition, was published in 2013 and was organized into 10 Knowledge Areas. Figure 16 provides a summary of the Knowledge Areas, where expert judgment is most frequently used according to the survey results.

There was no correlation between the expert judgment processes (seen in Figure 16) and the methods used (seen in Figure 15).

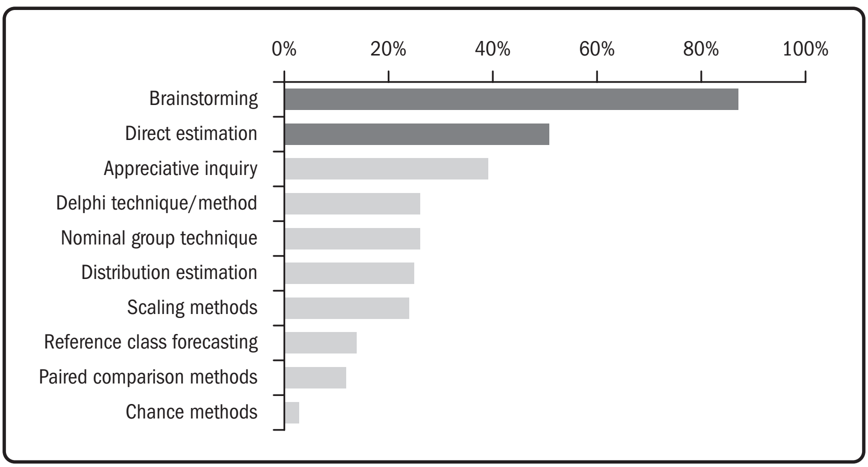


Figure 15 Methods used for eliciting expert judgment.

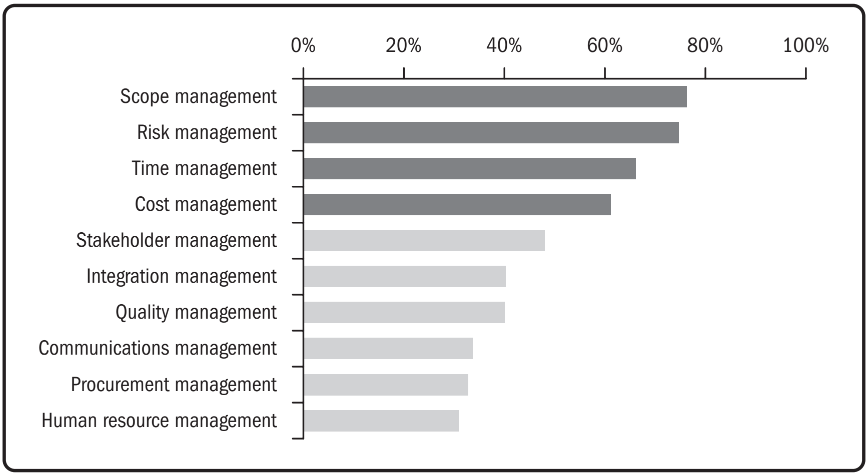


Figure 16 *PMBOK® Guide* processes where expert judgment is used.

3.4 Limitations

There were several limitations to this study. Some are common to survey methodology and others are unique to this study.

3.4.1 Population

Because the survey was posted to the Project Management Institute website on Survey Links and later sent to local PMI chapters, the population of the survey was taken to be the membership of the Project Management Institute or its local chapters. At nearly half a million members strong, the population was concentrated on members of the Project Management Institute. Even though PMI is the largest professional society and provider of certifications devoted to the advancement of project management, this study neglected the vast number of project management practitioners who are not affiliated with PMI, who may work independently of a professional membership organization, who may work with a different organization, or who may possess other credentials. Future studies may consider alternative populations to identify and confirm differences in practices.

3.4.2 Sample Randomness

The set of respondents was not a purely random sample in that the respondents were voluntary in nature and were not selected at random from the population. For a probabilistic data sample, one would need to target specific project management practitioners drawn randomly from the population. Without access to a master database of Project Management Institute members, this is not possible. Therefore, a chi-square test of goodness of fit was performed for each of the demographic variables provided in Table 9 to determine whether the region, experience, and industry of the sample was representative. The proportion of respondents from each region was not equally distributed in the population, χ^2 (dof = 5, N = 392) = 95.3, p > 0.005. The proportion of respondents from each level of experience was equally distributed in the population, χ^2 (dof = 3, N = 392) = 11.5, p > 0.005. The proportion of respondents from each level of industry sector was equally distributed in the population, χ^2 (dof = 6, N = 392) = 17.3, p > 0.005. Therefore, this sample was representative of the population at large across all but one of the demographic dimensions (i.e., region). This was largely because of the fact that the sample was significantly underrepresented for North America and overrepresented for Latin America and the Caribbean.

3.4.3 Sample Size

Despite 449 responses (of which 382 were complete), the sample represents less than one tenth of 1% of the population. The sample size was sufficiently large to analyze, but in the case of certain subsets (e.g., specific industry sectors or regions), the sample may have been too small for fine analysis. Therefore, most of the findings will focus on the entire response rather than examining responses of specific demographic elements. This was the scope of the study, and the analysis indicated that there were not statistically significant differences between subgroups (when size permitted this analysis).

Furthermore, because this was a pilot study (since a similar practice survey of expert judgment in project management has not been published), it is not possible to ascertain the actual effect size in order to conduct a formal power analysis. However, if we make some assumptions about anticipated effect based upon knowledge of the state of the practice of expert judgment in project management, we can perform a post hoc power analysis. To demonstrate, we will examine the portion of time a structured expert judgment process is used (see Figure 13). In practice, we would conservatively expect project management professionals to use such a process “more often than not” (or roughly 65% of the time), which would yield a mean response of “often” on the survey scale. In reality, respondents indicated that, on average, they used a structured process “sometimes” (or about 25% of the time, with a standard deviation of about 30%). If we assume the standard significance probability of 5% and set statistical power at the recommended level of 0.8 (Cohen, 1992), our post hoc power analysis would indicate that such a sample size ($n = 382$) would yield a power approach 1.0, meaning that it would correctly reject the null hypothesis when it was false.

3.4.4 Self-Reported Data

Based on its nature, this survey relied on the self-reporting of respondents regarding their practices. Though many have noted that self-report data is inherently biased and validity suffers as a result, there is an increasing body of evidence to support the use of self-report data (Chan, 2009; Specter, 2006). Also, although the interpretation of the terms of the survey may have varied between industries, regions, or applications, efforts were made to standardize the terminology and anchor it to the project management body of knowledge and lexicon. Additionally, the survey was field-tested using a think-aloud protocol to identify interpretation issues. Adjustments were made to help alleviate the interpretation issues.

3.4.5 Scope of Study

This survey was descriptive in nature and was focused on identifying the current state of the practice of expert judgment elicitation in project management. Explanatory research, which attempts to explain the reasons behind certain phenomena, was not conducted in this study. This may be a topic of future investigation.

3.5 Findings and Implications

There were several key findings in this descriptive survey that helped to shed light on the state of the practice:

- Expert judgment is widely used in project management. Over 95% of respondents indicated using expert judgment for the projects they manage.
- In most instances, expert judgment is conducted in an ad hoc manner. Almost three quarters of respondents indicated that they used a predefined structured process sometimes, rarely, or never (i.e., less than 40% of the time).
- Most practitioners work in organizations that do not have specific policies about using expert judgment. Only one in five respondents indicated he or she worked in an organization with written guidance for expert judgment elicitation.
- Most (57%) practitioners will use the written guidance for conducting expert judgment elicitation when it is available. Therefore, currently, only one in 10 project management practitioners uses written guidance to elicit expert judgment.
- There is a moderate correlation between the presence of written policy on expert judgment elicitation and practitioner use of a predefined structured process.
- Brainstorming, a free-form technique for generating ideas, is by far the most commonly used expert

judgment tool/technique. Eighty-seven percent of respondents indicated using brainstorming for conducting expert elicitation.

- Practitioners use expert judgment most frequently during scope, time, cost, and risk management processes. Respectively, 76%, 66%, 61%, and 75% of respondents reported using expert judgement in these processes (as compared to less than 50% reporting using expert judgment in the other process groups).

These findings strongly suggest that there is an opportunity to improve the state of the practice when it comes to using expert judgment in project management.

CHAPTER 4

Closing the Gap

The original intent of this phase was to conduct a series of experiments designed to match specific expert judgment elicitation methods to individual *PMBOK® Guide* project management processes. However, given dozens of processes and dozens of elicitation methods, this endeavor proved to be too cumbersome and would have produced a set of prescriptions rather than a descriptive means of identifying the appropriate elicitation method(s) for a particular context and purpose. Additionally, through the literature review, a categorization of elicitation methods as either generative or evaluative was presented, making the matching exercise unnecessary.

Instead, by examining the theory and practices, several gaps were uncovered. One of the foremost gaps, which occurs early in the planning stages of the expert judgment elicitation process, was the identification of requisite expertise for the selection of experts. Given this gap, the question becomes "How might we best select experts to provide us with informative judgments to allow for proper project management?" In an effort to address this theory-practice gap, two experiments were conducted. The first examined critical thinking as a potential way of selecting experts. The second experiment examined cognitive reflection and fluency as means for selecting experts.

4.1 Critical Thinking Experiment

As previously discussed, one of the principal biases experienced during expert elicitation is overconfidence. A wide variety of studies has been designed to identify which moderators may help calibrate expert confidence: for example, the impact of the availability of information (e.g., Oskamp, 1965; Tsai, Klayman, & Hastie, 2008), the impact of question difficulty (Lichtenstein et al., 1981), and the impact of feedback (e.g., Bolger & Önkal-Atay, 2004; Dawes, 1994; Lichtenstein & Fischhoff, 1977; Rauhut & Lorenz, 2010). One particularly interesting study indicated that experts' performance (and also confidence) was influenced by their cognitive styles (Tetlock, 2005). Borrowing from the idea of cognition, this study explored the notion that critical thinking ability may be an important determining factor in the selection of experts.

4.1.1 Participants

Participants were 86 undergraduate students, 45 juniors enrolled in an operations management course and 41 seniors enrolled in a capstone consulting course. The two groups represent two entire undergraduate cohorts of the business major at a selective, small public college located in New England. The population is mathematically sophisticated. All students take at least three mathematics courses, including a course in probability and statistics. The 25th and 75th percentiles of the student body scored between 590 and 650 on the mathematics portion of the SAT standardized examination. Given that all of the students are ultimately employed by a single organization in a single industry and that all the students participate in training and education (as well as 10 weeks of cooperative learning each summer) directly related to their industry, these participants were considered not just students but also apprentice experts.

4.1.2 Protocol

Because the experiment was intended to determine the correlation between critical thinking and overconfidence, two instruments were administered: 1) the Watson-Glaser Critical Thinking Appraisal (WGCTA), and 2) an eight-question expert elicitation instrument with several questions from the domain of expertise of the apprentice experts.

The first instrument, the WGCTA, was used to measure critical thinking. The WGCTA consists of 80 multiple-choice items divided into five subtests of 16 items each. The subtests were designed to measure the following aspects of critical thinking (Pascarella, 1989; Watson, 1980):

- Inference (discriminating inferences and their degrees of truth or falsity)
- Recognition of assumptions (recognizing unstated assumptions from given statements)
- Deduction (deciding if conclusions follow the information provided)
- Interpretation (deciding if conclusions are correct about the data by weighing evidence given)
- Evaluation of arguments (determining which arguments are strong or weak, as well as relevant or irrelevant)

Because the second instrument tested the experts' performance on evaluative expert judgment questions (and measured by their degree of overconfidence), it might be suggested that we might like to examine their deductive abilities (i.e., arriving at specific judgments using general theories and knowledge) or their inductive/inferential abilities (i.e., translating specific observations in general judgments). However, despite the fact that there are separate subtests, the WGCTA should be considered as a composite measure (Bernard et al., 2008) that has sufficient validity and reliability as a composite (Gadzella et al., 2006).

The second instrument was an expert elicitation questionnaire that consisted of eight almanac-type questions (see Appendix B for the complete instrument). Its design was based upon the format of a method comparison study conducted by the Australian Centre of Excellence in Risk Analysis (ACERA) that examined the influence of question design (i.e., three- versus four-point estimation) on expert overconfidence (Speirs-Bridge et al., 2010).

Table 11 Expert elicitation questions.

Group A	Group B
• Length of U.S. coastline (in miles) • Vehicles crossing U.S./Mexico border in 2010 • Amount of commodities carried through Louisiana ports in 2007 (in tons) • Odds of person living in the United States being struck by lightning in lifetime	• Water surface area of U.S. great lakes (in square miles) • Average number of passengers flying domestically in the United States in 2010 • Operating revenues for U.S. passenger ferries in 2009 (in $U.S.) • Probability of getting a straight flush in five-card poker

Table 11 describes the topics contained in the eight questions. The questions were broken out into two groups (i.e., groups A and B) to experimentally examine if order of questions or order of elicitation methods had any effect on judgments.

Table 12 shows the four versions of this instrument devised to control for the order of questions and to test the mode of elicitation (Appendix B contains Form A4B3).

4.1.3 Method

The first instrument, the WGCTA, was administered by the institutional research function at the college during a predetermined testing period using a previously established protocol, as a part of the institution's learning assessment program. The WGCTA form B is administered to seniors each year (and form A is administered to freshmen each year).

The second instrument was administered as part of a class in expert elicitation. The juniors were administered two instruments

Table 12 Four versions of expert judgment protocol.

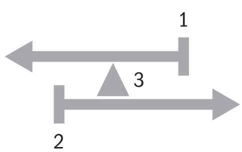
		Order of Elicitation Method	
		3-point then 4-point	4-point then 3-point
Order of Question Groups	Group B then Group A	Form B3A4 (n = 22)	Form B4A3 (n = 21)
	Group A then Group B	Form A3B4 (n = 22)	Form A4B3 (n = 21)

on separate days during the 2011 fall semester. The seniors were administered the same two instruments on separate days in the 2012 spring semester. In addition to a verbal explanation on how to complete three- and four-point estimates (with examples provided), each elicitation worksheet had instructions for each elicitation procedure (see Figure 17).

Each participant was given one of the four versions of the instrument and provided 15 minutes to complete the elicitation instrument. Participants were instructed to provide their student identification numbers so that their judgments could be matched with their critical thinking scores. Participants were assured anonymity. Following this elicitation exercise, the class then went on to examine expert elicitation, examples from across a variety of contexts, and the research (including that cited in this paper regarding overconfidence and interval estimation).

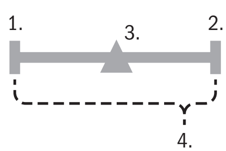
Once all the responses were collected, the data were compiled. In order to facilitate comparison of the three-point results and four-point results, whenever a four-point interval had other than an 80% interval, a transformation was conducted to generate an equivalent 80% interval for purposes of comparison. The

3-point estimation
INSTRUCTIONS:



For each quantity below, please provide the following <u>three</u> estimates:
1. I am 90% confident the true value will be ***less than*** _____.
2. I am 90% confident the true value will be ***greater than*** _____.
3. Realistically, the ***most likely value*** is _____.

4-point estimation
INSTRUCTIONS:



For each quantity below, please provide the following <u>four</u> estimates:
1. What do you think the ***minimum*** value could practically be?
2. What do you think the ***maximum*** value could practically be?
3. Realistically, what is the ***most likely*** value?
4. How ***confident*** are you the interval you created will capture the true value? Please enter a number between 50% and 100%.

Figure 17 Interval elicitation procedure.

transformation method prescribed in Speirs-Bridge et al. (2010) was used to facilitate comparison.

4.1.4 Results

It was hypothesized that critical thinking would be an effective moderator in reducing overconfidence. It was proposed that individuals with higher critical thinking scores would examine the limitations of their knowledge and as a result would adjust their judgments accordingly to reflect their uncertainty (i.e., they would increase the intervals by creating a larger spread between their minimum and maximum estimates in order to capture the true values). However, there was insufficient evidence in this experiment to support this notion. Based on the ANOVA, only 10.3% of the variation in overconfidence could be explained by critical thinking. Thus, critical thinking ability as measured by the WGCTA lacked predictive power to suggest that overconfidence would be reduced. Figure 18 illustrates this with a plot of participants' overconfidence as it related to critical thinking (each plot point representing one or several respondents) and the linear trend line best fit to the data.

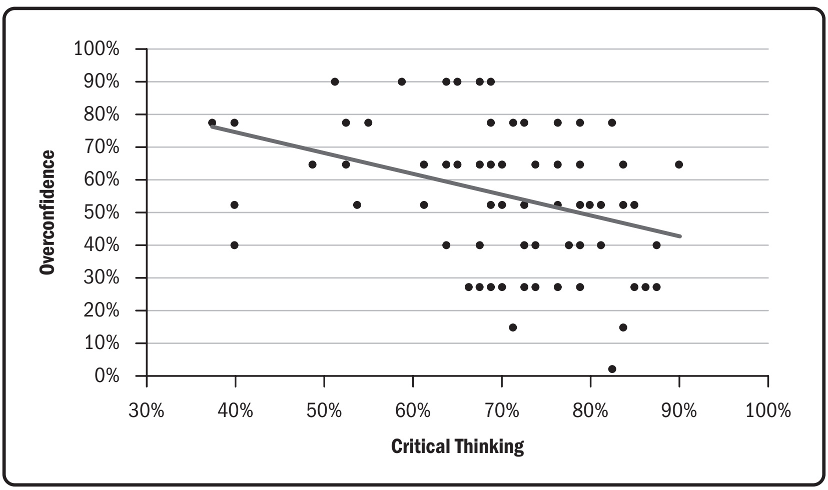


Figure 18 Critical thinking versus overconfidence.

Furthermore, when looking at the five subscale dimensions of critical thinking contained in the WGCTA, only *evaluation of arguments* was statistically significant at the 95% level.

Despite this inconclusive result, the experiment demonstrated an average reduction in overconfidence of 7% comparing the four-point to the three-point estimation methods (compared to a 28% reduction when using the four-point method and a 12% reduction when using the three-point method in the ACERA study). These results are consistent with previous studies. However, the fact that overconfidence was higher in this study may be a result of the fact that students were estimating almanac questions and were not actually experts estimating values within their fields of specialization.

This study also examined how order of magnitude related to interval estimates. The true values of the group A questions had order of magnitude 10^5, 10^8, 10^5, and 10^{-5} respectively. The true values of group B questions had order of magnitude 10^5, 10^8, 10^8, and 10^{-5} respectively. The responses to the questions were sorted based upon order of magnitude (where $M = 10^8$, $K = 10^5$, and $\mu = 10^{-5}$).

Responses were then categorized as follows:

- *Underestimate:* The assigned interval was nearer origin than true value.
- *Accurate:* The assigned interval captured true value.
- *Overestimate:* The assigned interval was farther from origin than true value.

Figure 19 illustrates the results when the response categories are sorted according to the order of magnitude. It also includes all responses because there was no statistical difference between the order of question groups and the order of elicitation methods (see Table 12 for the various potential combinations). In general, about one third of the responses were accurate regardless of the order of magnitude of the questions. However, judges tended to underestimate the questions with the highest order of magnitude (M) and overestimate those with the lowest order of

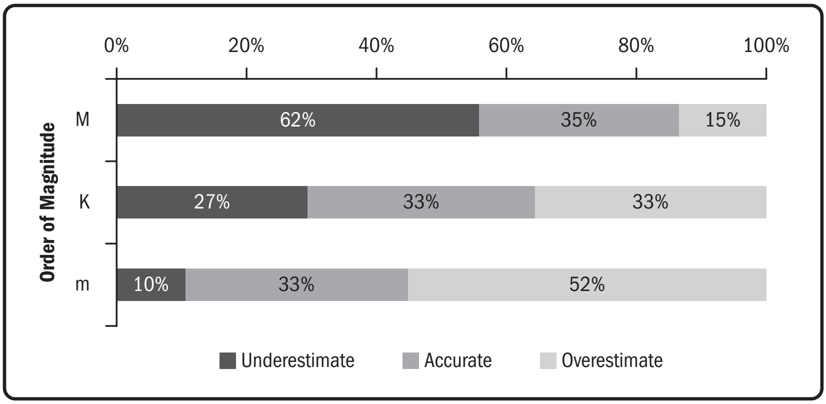


Figure 19 Under- versus overestimation by order of magnitude.

magnitude (μ). Those questions in the middle order of magnitude (K) were equally likely to be overestimated or underestimated. Although these last two findings do not directly relate to expert selection, they are worth noting because they may have an impact on planning the elicitation (e.g., choosing an appropriate elicitation method) and training the experts (about magnitude effects).

4.2 Numeracy and Fluency Experiment

Borrowing from Fasolo and Bana e Costa's (2014) work on elicitation preferences based upon numeracy and fluency, this study explores the notion that fluent experts will perform better on generative expert judgment tasks and numerate experts will perform better on evaluative expert judgment tasks.

4.2.1 Participants

The study was originally conducted on 37 participants who were project management professionals employed by a governmental agency, but the permission to use the data for this study was later reneged based upon an internal review. Despite this setback, the study was conducted a second time on 39 participants, and as in the prior experiment, the participants were undergraduate students (seniors and juniors) enrolled in a logistics course at a public special mission college located in New England. The student population has the following mean SAT scores: 570 math, 520 critical reading, and 520 writing (COLLEGEdata.com, 2015). The accepted profile of the students for the entire school had overall SAT scores ranging from 1,400 to 1,680 (AcceptanceRate.com, 2015). Though it might have been interesting to see if SAT math scores moderated numeracy and evaluative expert elicitation tasks (and also if SAT critical reading and/or writing scores moderated literacy and generative expert elicitation tasks), individualized SAT information was not available and, therefore, tests for determining numeracy and fluency were used instead.

4.2.2 Protocol

Because this experiment was designed to evaluate both the impact of expert numeracy on evaluative expert judgment elicitation tasks and the impact of expert fluency on generative expert judgment elicitation tasks, four instruments were involved.

Numeracy Assessment: To evaluate participant numeracy, a hybrid numeracy scale was developed to include questions involving a "riskless" context numeracy scale (Fasolo & Bana e Costa, 2014), as well as the general numeracy scale that involved elements of risk and uncertainty (Lipkus et al., 2001; Peters et al., 2006). The expanded portion of the risk context numeracy scale was focused on converying risk information to patients in a healthcare setting (Woloshin, Schwartz, Moncour, Gabriel, & Tosteson, 2001; Zikmund-Fisher, Smith, Ubel, & Fagerlin, 2007) and, as a result of its unique context, was adapted for inclusion in this study. A copy of this instrument has been provided in Appendix C. Each participant was assigned a numeracy score based upon the total number of correct answers.

Evaluative Judgment Assessment: To evaluate participant performance on an evaluative expert judgment elicitation task, the instrument used was the same expert elicitation questionnaire (that consisted of eight almanac-type questions) used in the previous experiment (see Table 11 for the topics covered and Appendix B for the complete instrument). Participants' performance on this instrument was measured by their degree of under- or overconfidence. As is common in similar studies, overconfidence is measured by subtracting the participant's *hit rate* (i.e., the proportion of accurate estimates) from the reference confidence intervals. For example, if expert A is asked to estimate the 10th and 90th percentile intervals for 10 quantities of interest and that expert's intervals only capture five of the true values, the expert would exhibit 30% overconfidence (80% − 50% = 30%). On the other hand, if expert B captures all 10 values within his or her interval estimates, the expert would exhibit 20% *under*confidence (80% − 100% = −20%).

Fluency Assessment: Participant fluency was measured using the Controlled Oral Word Association Test (Spreen & Strauss, 1998). Participants were given a letter from the alphabet and instructed to write down all the words that they could think of that began with that letter in a three-minute period. This was repeated three times for the letters A, F, and L, as is customary in the administration of this test. Each participant was assigned a fluency score for the total number of words he or she generated in the nine-minute period.

Generative Judgment Assessment: To evaluate participant performance on a generative expert judgment elicitation task, participants were provided with a short case study from the Project Management Institute's online case study library. In this instance, it was the April 2014 case study entitled "Project Management Helps Create World's Longest Natural Gas Pipeline," which involved the installation of a 9,000-kilometer-long pipeline providing power to 500 million Chinese residents and the city of Hong Kong (Project Management Institute, 2014). In a 10-minute period, participants were asked to produce as many risks or hazards to the project as they could identify. Though performance on this task might be moderated by project management knowledge and experience, the participants selected had minimal project management training and experience. The specific case was chosen because it provided a subject and context that would be familiar to the participants based upon their chosen major and the mission of the institution where they were studying. Further, this case study was deemed appropriate because it presented a robust risk space and contained political, economic, social, and technical risks. Participants' performance on this task was measured by the total number of distinct risks/hazards they generated for the scenario and project presented in the case study. The idea was that when trying to identify risks (as in the *PMBOK® Guide* Project Risk Management Process Group 11.2), quantity is as important as quality. The more risks you can identify, the more likely you are to capture a full range of potential risks that might be experienced in a project.

4.2.3 Methods

Participants were administered the protocol (consisting of all four instruments) during one of three different sessions administered in March 2015. The participants were informed that the experiment consisted of four sections (and each section was briefly described). The participants were given the first few moments to answer a few questions about their previous project management training and experience. Next, the participants were instructed on and administered the timed fluency portion of the experiment. They were then given the numeracy portion of the experiment. They were given five minutes to complete the five-item numerical scale, without a calculator. Next, the participants were trained in three-point and four-point estimation and instructed on how to complete the expert elicitation worksheet (Appendix B). Finally, the participants were provided the case study to read. After reading the case study, participants were instructed to write down as many potential risks and hazards as they could think of for the project in the case study.

4.2.4 Results

All participants had minimal formal project management training (mean of 5–10 hours) and minimal project management experience (all less than one year of experience). Participant numeracy was on the interquartile range of the spectrum (mean score of 53.1%, median of 5 correct out of 10, range 2 to 10) and the reliability of the numeracy test was adequate (Cronbach's alpha = 0.65). Participant fluency was also on the interquartile range of the spectrum (mean score of 66, median of 69, range of 45 to 129) and the reliability of the fluency test was strong (Cronbach's alpha = 0.82). Because the distributions were skewed, median splits were performed on both the numeracy and the fluency measures. The sample size was much too small to allow for quartile or other splits.

The analysis of numeracy as a moderator compared low numeracy (2, 3, 4, or 5 correct) to high numeracy (6, 7, 8, 9, or 10 correct). The independent variable in this analysis was the amount of overconfidence calculated from the participants' estimates on the elicitation worksheet. When the sample was split into high and low numeracy groups, the results indicated that less numerate participants had higher degrees of overconfidence on the evaluative expert judgment task. The difference was statistically significant, $t(37) = 2.32$, $p < 0.05$. Figure 20 illustrates this result for the evaluative expert elicitation task. Simply put, experts with

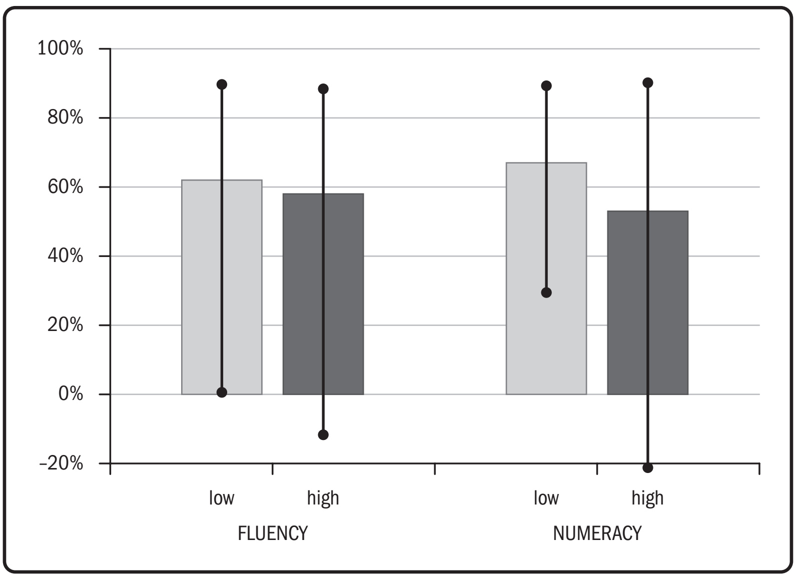


Figure 20 Effect of numeracy and fluency on evaluative expert elicitation tasks.

higher numeracy made better estimates than did the experts with lower numeracy. Or put another way, experts with lesser numeracy tended to be more overconfident in their estimates. A similar non-significant difference was observed between the low fluency and high fluency groups; however, there was a moderate correlation between fluency and numeracy ($r = 24.1\%$).

Similarly, a median split was performed on the fluency measure. The analysis compared low fluency (45 to 69 words) to high fluency (70 to 129 words). In this analysis, the independent variable was the number of distinct risks identified by the participants regarding the case study provided. When comparing fluency levels, there was a statistically significant difference between the number of risks generated by the high fluency group and the low fluency group: $t(37) = 2.94$, $p < 0.01$. Simply put, experts with higher levels of fluency were able to generate more potential alternatives in the task. As in the previous analysis, there was a difference in numeracy levels as well, but it was not as pronounced and was not significant. These results are illustrated in Figure 21.

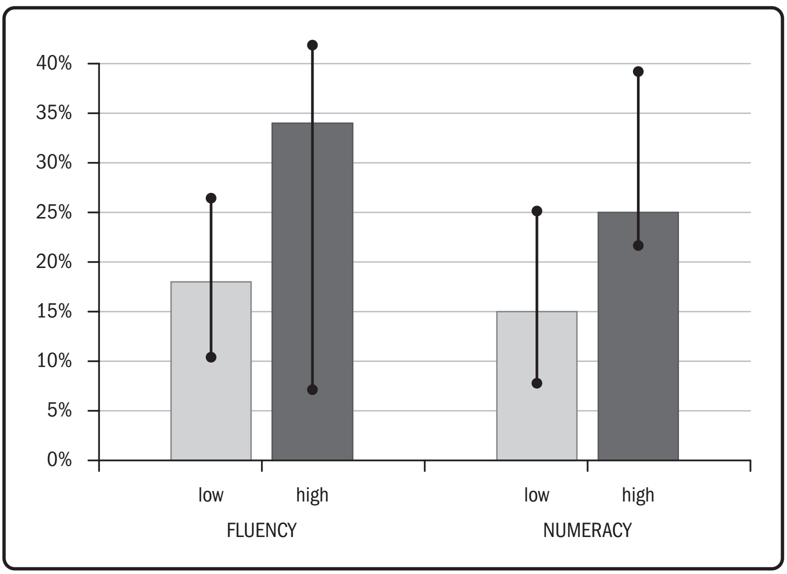


Figure 21 Effects of numeracy on evaluative and generative expert elicitation tasks.

4.3 Findings and Implications

Although limited in scope and subject to substantial limitations, these experiments provided some direction toward the task of selecting experts:

- Critical thinking was not demonstrated to be a significant moderator for reducing overconfidence in evaluative judgments. If this is to be considered a method for expert selection, further study will be required.
- Simple fluency tests appear to be useful in identifying which experts will perform best on generative tasks.
- Simple numeracy tests appear to be useful in identifying which experts will perform best on evaluative tasks.

These findings would suggest that there are (or will be) effective means for identifying which experts possess the requisite expertise for the particular form of expert judgment information needed.

CHAPTER 5

Discussion

5.1 Summary

Phase 1 helped us identify the current state of the practice of expert judgment within project management. Phases 2 and 3 provided us insight about how to improve the practice of expert judgment within project management using the state of the art/science found outside project management, and thereby narrow the theory-practice gap. From the perspective of Kerzner's project management maturity model (2011) (shown in Figure 22), this study should help provide a pathway for developing basic knowledge about expert judgment and defining the process of expert judgment in an effort to move from level 1 through level 2

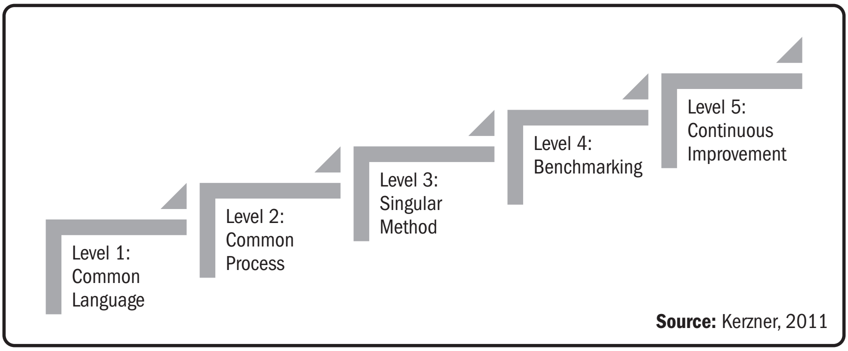


Figure 22 Kerzner's project management maturity model.

and into level 3, and ultimately beyond with time. In sum, by introducing language of basic knowledge and process definitions about expert judgment, it is expected that the practice of using expert judgment in project management would move from an immature level 1 stage to a more sophisticated level 3 stage of project management maturity.

5.2 Key Findings

The following are some general findings and conclusions that have resulted from this study.

5.2.1 State of the Art/Science Is Established and Growing

Based on the literature review conducted in phase 1 of this study, it is clearly apparent that other disciplines are more advanced than is project management when it comes to expert judgment elicitation methods (particularly the evaluative methods). Even though the scope of the review was limited to a one-year period, there were more than 100 relevant articles on the subject of expert judgment (see Figure 3), of which many have been captured in this report. Upon reviewing Chapter 2 and the scope of the references, it is clear to see there is a robust body of knowledge outside project management in regard to expert judgment elicitation. A more comprehensive review of the literature (spanning a greater period of time) would yield even more opportunities for improvement.

5.2.2 State of the Practice in Project Management Is Informal and Emergent

Based upon the descriptive survey of phase 2 of this study, it is clear that the state of the practice of expert judgment within project management resides within level 1 (or optimistically level 2) of Kerzner's project management maturity model (seen in Figure 22). It is based on ad hoc (and likely inconsistent) processes for conducting expert judgment, typically conducted without the benefit of written guidance. Presented in contrast to the findings of Chapter 2, there is a clear gap between the practice and the theory, and as a result, a significant opportunity exists to improve the practice.

5.2.3 Expert Judgment Elicitation in Project Management Can Mature

The experiments of phase 3 of this study provide evidence that additional informative research and evidence can be applied to mature and standardize the practice of expert judgment elicitation in project management to eventually become a repeatable, well-defined, and structured tool/technique within the toolkit of project management practitioners everywhere.

In that direction, the next suggestion provides a series of suggested practices to improve the practice of expert judgment in project management.

5.3 Suggested Practices

5.3.1 Use a Generic Process

Because much of the practice of project management employs a variety of process models, such a model should be used when using expert judgment as a tool/technique. Currently, only one in 10 project management practitioners uses written guidance or policy for expert judgment elicitation. The presence of written guidance is strongly correlated to the frequency of practitioner usage of predefined structured processes, which are known to help alleviate judgmental biases and, thus, improve the accuracy of estimates. There is a variety of existing processes for eliciting expert judgment; several have been provided in Chapter 2. Most processes are sufficiently malleable to allow adjustment for individual circumstances. We suggest adopting or creating a generic process and providing policy on how to conduct expert judgment elicitation. The generic seven-step process presented in Chapter 2 will now be used to make further suggestions on how to elicit expert judgment.

5.3.2 Frame the Problem

Clearly identifying the specific information you need to obtain through expert judgment is critical. Solving the wrong problem perfectly is no better than solving the right problem imperfectly. As a start, if you are using the *PMBOK® Guide* set of processes, you may be interested in looking up the indicated output(s) for the specific process where you intend to employ expert judgment as tool/technique (see Table 4). Once you have determined the form of the information sought, you will be able to identify whether the expert judgment task will be evaluative or generative in nature. Recall that evaluative methods result in expert judgment that evaluates (or otherwise estimates, forecasts, predicts, or quantifies) desired information. Examples include cost or time estimates and risk probabilities or impacts. Generative methods result in expert judgment that generates lists or descriptions of desired information. Examples include activity lists, risk registers, and stakeholder lists. By naming your desired information as either evaluative or generative, you will be able to select an appropriate method.

5.3.3 Plan the Elicitation

In order to get the best judgments or estimates, you must select the most appropriate method from among the hundreds available. Some methods work best for evaluative tasks and others for generative tasks. In the *PMBOK® Guide*, the generative methods (although not classified that way) are grouped under the lists and descriptions of group creativity techniques, alternatives generation, and group decision-making techniques:

- Brainstorming
- Nominal group technique
- Mind mapping
- Affinity diagramming
- Delphi technique
- Lateral thinking
- Analysis of alternative

This report provides a more extensive list of available methods (particularly the ones for evaluative tasks) than does the *PMBOK® Guide*. Once you have selected a method that best suits your task, an expert judgment elicitation protocol should be designed. The design of the elicitation should include the following elements:

- Type of information sought (i.e., generative or evaluative)
- Specific information requiring expert judgment
- Method for eliciting expert judgment (e.g., see Tables 5 and 6)
- Mode for eliciting the expert judgment (e.g., interactive group, nominal group, individual)
- Type of expertise required and method of expert selection
- Form of pre-elicitation training (to inform about process and to debias)

Using these elements as a starting point, an elicitation protocol can be developed. It should be field-tested to determine face validity and to identify which areas require improvement or refinement.

5.3.4 Select Experts

Expert selection is paramount for expert judgment to be useful and informative. To determine how best to select experts, having just identified the type of information and method/mode of elicitation, you will now identify the requisite expertise. This might include the following:

- Expertise about scope and activities needed to complete a specific project
- Expertise about costing and estimates for work packages
- Expertise about potential risks posed by various phases of a specific project
- Expertise about stakeholders

Select four to eight experts (from inside the project team and outside of the organization to ensure a diversity of perspectives) who have both the necessary experience and knowledge, as well as the appropriate credentials.

5.3.5 Train Experts

Train the experts before the actual expert judgment elicitation. Share with them what information is being sought and why/how it will be used in managing the project. Provide experts with awareness of various biases, heuristics, and common pitfalls. Also give them practical means of dealing with these issues, often through practice. Demonstrate the elicitation protocol through practice problems.

5.3.6 Elicit Judgments Using Appropriate Methods

Actual elicitation requires familiarity with the expert judgment method. After determining if a generative or evaluative method is required, select a specific elicitation method. Note that although brainstorming is most frequently used, it is not appropriate for many expert judgment elicitation situations, including obtaining evaluative information. Do not over-rely on brainstorming or other ad hoc methods; instead, select the best method and read up on how it works. Then practice the method with normative experts to develop proficiency in administering the elicitation. After that, elicit the judgments.

5.3.7 Analyze Judgments and Combine (if Desired)

If possible, assess expert performance (e.g., through performance on seeded variables) to determine weights and which expert judgments should be included. However, do not arbitrarily omit expert judgment. For generative elicitations, behavioral aggregation of judgments (e.g., consensus) will be sufficient. For evaluative elicitations, simply average expert judgments to get an aggregate judgment unless you have a compelling reason to do otherwise (since the research has shown averaging to outperform other combination techniques in most cases).

5.3.8 Document Results and Communicate

Documentation is necessary for both historic purposes and communication purposes. The documentation of expert judgment should allow any reviewer to reconstruct the logic and outcomes of the expert elicitation. Documentation should occur at all seven steps of the generic protocol, including the following:

1. Problem Statement: A concise statement of the problem and list of information needed
2. Elicitation Plan: An expert judgment plan and a validated elicitation protocol (with instructions)
3. Expert Selection: A list of experts, their affiliations, and their curricula vitae
4. Expert Training: Lesson plan for training, copies of training materials, list of attendees, and questions/lessons learned for future improvement
5. Judgment Elicitation: Compilation of all judgments and rationale for why those judgments were made (and notes about the elicitation process)
6. Judgment Aggregation: A description of the method, documentation of any interaction, and the aggregated judgment
7. Elicitation Documentation: Using the above information, a comprehensive report should be developed to communicate the expert judgments and how they were developed

Appendix A

GLOBAL STUDY OF EXPERT JUDGMENT PRACTICES IN PROJECT MANAGEMENT

Invitation to Participate

You are cordially invited to participate in this survey of current *expert judgment* practices, a study sponsored by the Project Management Institute (PMI).

The purpose of this study is to develop a clear understanding of current expert judgment practices in project management. Your participation is greatly appreciated.

Consent to Participate

Your participation in this survey is voluntary. You have the right to stop it at any time or for any reason, without adverse consequences.

The information you provide us will be anonymous and confidential. Reported findings will be non-attributable. Data will be stored securely and will be aggregated for academic and research purposes and referred to in any publications that may result from this survey.

By clicking on the NEXT button below and starting this survey, you accept these terms.

Instructions for Completing the Survey

Please read carefully.

Once started, the average time to complete this survey is about 10 minutes.

This survey includes 14 questions.

Please answer all questions. Otherwise, your responses will be invalidated. Once you start the survey, complete all the questions and submit the survey to prevent loss of data.

Use the NEXT button to move forward through the survey. You may also return to previous responses using the PREVIOUS button.

Respondent & Organization Characterization

1. Which <u>best</u> describes your *primary job function*?
   - ❍ Project Management Consultant
   - ❍ Project Management Specialist/Team Member
   - ❍ Project Manager
   - ❍ Program/Portfolio Manager
   - ❍ Director of Project/Program Management Office
   - ❍ Other; please specify: ______________________

2. How many years of *project management experience* do you have?
   - ❍ Less than 1 year
   - ❍ 1 to 5 years
   - ❍ 6 to 15 years
   - ❍ More than 15 years

3. Do you have project management certification or credentials (e.g., PMP®, governmental certification, internal company-sponsored certification)?
   - ❍ Yes, please specify which: ______________________
   - ❍ No

4. Which *types of projects* do you primarily manage and/or participate in?

5. In which *country* is your office located (i.e., your primary work site)?

Expert Judgment Practices

According to *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)* – Fifth Edition, **expert judgment** is judgment provided based upon expertise in an application area, knowledge area, discipline, industry, etc., as appropriate for the activity being performed. Such expertise may be provided by any group or person with specialized knowledge, education, experience, or training.

6. Do you use expert judgment in the projects you manage or participate in?
   - ❍ Yes
   - ❍ No

7. Why do you *use* expert judgment? Check ALL that apply.
   - ❑ To obtain facts or figures about unknown quantities
   - ❑ To identify possible future events or activities
   - ❑ To describe possible future scenarios
   - ❑ Other; please specify: ____________________

8. In which *project management process* do you most frequently use expert judgment? Check ALL that apply.
   - ❑ Integration management (e.g., chartering, change control, and closeout)
   - ❑ Scope management (e.g., requirements, work breakdown, and scope control)
   - ❑ Time management (e.g., estimating durations, scheduling, and sequencing)
   - ❑ Cost management (e.g., estimating cost, budgeting, controlling cost)
   - ❑ Quality management (e.g., quality assurance and quality control)
   - ❑ Human resource management (e.g., staffing, developing/managing team)
   - ❑ Communications management (e.g., managing communications)
   - ❑ Risk management (e.g., identifying hazards, estimating risk, analyzing risk)
   - ❑ Procurement management (e.g., performing procurements)
   - ❑ Stakeholder management (e.g., identifying stakeholders, engaging stakeholders)

9. Does your organization have *written guidance* on how to elicit expert judgments?
   - ❍ Yes, and it is typically followed
   - ❍ Yes, but it is rarely followed
   - ❍ No

10. How often do you follow a predefined, *structured process* for eliciting expert judgments?
    - ❍ Never (0%)
    - ❍ Rarely (1–10%)
    - ❍ Sometimes (11–40%)
    - ❍ Often (41–70%)
    - ❍ Typically (71–99%)
    - ❍ Always (100%)

11. Which of the following *expert judgment* elicitation tools and techniques do you most frequently use? Check ALL that apply.
    - ❑ Appreciative inquiry
    - ❑ Brainstorming
    - ❑ Chance methods (lottery wheels, odds ratios, etc.)
    - ❑ Delphi technique/method
    - ❑ Direct estimation (i.e., single-point estimation)
    - ❑ Distribution estimation (multipoint, quantile estimation, etc.)
    - ❑ Nominal group technique
    - ❑ Paired comparison (Analytic Hierarchy Process, Bradley-Terry model, etc.)
    - ❑ Scaling methods (discrete/continuous scales, order ranking)
    - ❑ Reference class forecasting
    - ❑ Other; please specify: ______________________________

12. How do you *select* your experts?

____

____

____

13. How do you *combine* the judgments of multiple experts?
    - ❍ Mathematical (straight or weighted average, performance weighting, etc.)
    - ❍ Consensus
    - ❍ Other; please specify: ____________________

14. Describe any effective practice(s) that you use for eliciting expert judgment.

____

____

Thank You—Survey Complete

Thank you for your participation in this survey.

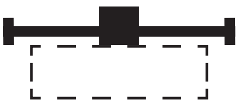
At the end of this study, you will be eligible to access and download a copy of the Executive Report containing a summary of the results, main research findings, and managerial implications. The report will be available online at PMI.org within 60 days after the survey completion and data analysis.

For questions, please contact the principal investigator, Paul Szwed, at pszwed@maritime.edu.

Appendix B

Expert Elicitation Worksheet

INSTRUCTIONS: For each quantity below, please provide the following <u>four</u> estimates:



1. What do you think the ***minimum*** value could practically be?
2. What do you think the ***maximum*** value could practically be?
3. Realistically, what is the ***most likely*** value?
4. How ***confident*** are you the interval you created will capture the true value? Please enter a number between 50% and 100%.

What is the ***length of the U.S. coastline***, including Alaska and Hawaii and all territories (in miles)?

1. Practical Maximum: ______________________ miles
2. Practical Minimum: ______________________ miles
3. Most Likely Value: ______________________ miles
4. Confidence: _____ % (Please enter a number between 50% and 100%.)

What was the total number of ***vehicles crossing the United States/Mexico border in 2010***?

1. Practical Maximum: ______________________ vehicles
2. Practical Minimum: ______________________ vehicles

3. Most Likely Value: ____________________ vehicles
4. Confidence: _____ % (Please enter a number between 50% and 100%.)

What was the number of ***tons of commodities carried by vessels that flowed through ports in Louisiana in 2007***?

1. Practical Maximum: ____________________ tons
2. Practical Minimum: ____________________ tons
3. Most Likely Value: ____________________ tons
4. Confidence: _____ % (Please enter a number between 50% and 100%.)

What are the ***odds that a person living in the United States will be struck by lightning in his or her lifetime***?

1. Practical Maximum: one in ____________________
2. Practical Minimum: one in ____________________
3. Most Likely Value: one in ____________________
4. Confidence: _____ % (Please enter a number between 50% and 100%.)

Expert Elicitation Worksheet

INSTRUCTIONS: For each quantity below, please provide the following three estimates:

1. I am 90% confident the true value will be ***less than*** ______.
2. I am 90% confident the true value will be ***greater than*** ______.
3. Realistically, the ***most likely value*** is ______.

What is the ***water surface area of the Great Lakes (includes Lakes Superior, Michigan, Huron, Erie, and Ontario) (in square miles)***?

1. I am 90% confident the true value will be less than ______________ square miles.
2. I am 90% confident the true value will be greater than ______________ square miles.
3. I estimate the most likely value is ______________ square miles.

What was the average monthly number of ***passengers flying domestically in the United States in 2010***?

1. I am 90% confident the true value will be less than ______________ passengers.
2. I am 90% confident the true value will be greater than ______________ passengers.
3. I estimate the most likely value is ______________ passengers.

What is the total ***passenger operating revenue (in $U.S.) for ferryboats operating in the United States in 2009 (excluding international, rural, rural interstate, and urban park ferries)***?

1. I am 90% confident the true value will be less than ______________ $U.S.
2. I am 90% confident the true value will be greater than ______________ $U.S.
3. I estimate the most likely value is ______________ $U.S.

What is the probability of getting a straight flush in five-card poker?

(For example, 2♣, 3♣, 4♣, 5♣, 6♣ is a straight flush.)

1. I am 90% confident the true value will be less than ______________.
2. I am 90% confident the true value will be greater than ______________.
3. I estimate the most likely value is ______________.

Appendix C

Numeracy Scale—General

1. Imagine that we rolled a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even (2, 4, or 6)?
2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize is 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each bought a single ticket to BIG BUCKS?
3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to the ACME PUBLISHING SWEEPSTAKES win a car?

Numeracy Scale—Extended (Adapted)

4. Which of the following numbers represents the greatest risk?
   a. 1 in 100
   b. 1 in 1,000
   c. 1 in 10
5. If Person A's risk of getting a disease is 1% in 10 years, and person B's risk is double that of A's, what is B's risk?
6. If Person A's chance of getting a disease is 1 in 100 in 10 years, and person B's risk is double that of A's, what is B's risk?
7. If the chance of getting a prize is 10%, how many people in 1,000 would be expected to win the prize?

8. If the chance of getting a prize is 1 out of 50, this would be the same as having a ______% chance of getting the disease.
9. The chance of getting a viral infection is 0.0005. Out of 10,000 people, about how many of them are expected to get infected?

References

AcceptanceRate.com. (2015, June 15). *Acceptance rate & average GPA*. Retrieved from http://www.acceptancerate.com

Adams, J. L. (1986). *Conceptual blockbusting.* Cambridge, MA: Perseus Books.

Adelman, L., & Bresnick, T. A. (1992). Examining the effect of information sequence on Patriot air defense officers. *Organizational Behavior and Human Decision Processes, 53*(2), 204–228.

Adelman, L., Tollcott, M. A., & Bresnick, T. A. (1993). Examing the effect of information sequence on expert judgment. *Organizational Behavior and Human Decision Processes, 56*(3), 348–369.

Ahlemann, F., Arbi, F. E., Kaiser, M. G., & Heck, A. (2013). A process framework for theoretically grounded prescriptive research in the project management field. *International Journal of Project Management, 31*(1), 43–56.

Alahuhta, P., Nordbäck, E., Sivunen, A., & Surakka, T. (2014). Fostering team creativity in virtual worlds. *Journal for Virtual Worlds Research, 7*(3).

Aliakabargolkar, A., & Crawley, E. F. (2014). A Delphi-based framework for systems architecting of in-orbit exploration infrastructure for human exploration beyond low earth orbit. *Acta Astronautica, 94*(1), 17–33.

Aloysius, J. A., Davis, F. D., Wilson, D. D., Taylor, A. R., & Kottwmann, J. E. (2006). User acceptance of multi-criteria judgments appeal differently to decision makers. *European Journal of Operational Research, 169*(1), 273–285.

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, &

A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 294–305). Cambridge, UK: Cambridge University Press.

Anderson, J. R., & Sunder, S. (1995). Professional traders as intuitive Bayesians. *Organizational Behavior and Human Decision Processes, 64*(2), 185–202.

Arkes, H. R., Christianson, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behaviour and Human Decision Processes, 39*, 133–144.

Armstrong, J. S. (2011). *Principles of forecasting: A handbook for researchers and practitioners.* Berlin, Germany: Heidelberg.

Ayyub, B. M. (2001). *Elicitation of expert opinions for uncertainty and risks.* Boca Raton, FL: CRC Press.

Baker, E., Bosetti, V., Jenni, K. E., & Ricci, E. C. (2014). *Facing the experts: Survey mode and expert elicitation.* Milan, Italy: Fandazione Eni Enrico Mattei.

Bane e Costa, C. A., De Corte, J. M., & Vansnick, J. C. (2012). MACBETH. *International Journal of Information Technology and Decision Making, 11*, 359–387.

Barnum, C. M. (2010). *Usability testing essentials: Ready, set, . . . test!* Burlington, MA: Morgan Kaufmann.

Barron, F. H. (1992). Selecting a best multiattribute alternative with partial information about attribute weight. *Acta Psychology, 80*(1–3), 91–103.

Beaty, B. E., Benedek, M., Wilkins, R. W., Jauk, E., Silvia, P. J., & Neubauer, A. C. (2014). Creativity and the default network: A functional connectivity analysis of the creative brain at rest. *Neuropsychologia, 64*, 92–98.

Bernard, R. M., Zhang, D., Abrami, P. C., Sicoly, F., Borokhovski, E., & Surkes, M. A. (2008). Exploring the structure of the Watson-Glaser Critical Thinking Appraisal: One scale or many? *Thinking Skills and Creativity, 3*(1), 15–22.

Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *International Journal of Forecasting, 20*(1), 29–39.

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision Support Systems, 11*(1), 1–24.

Bottomley, P. A., Doyle, J. R., & Green, R. H. (2000). Testing the reliability of weight elicitation methods: Direct rating versus point allocation. *Journal of Marketing Research, 37*(4), 508–513.

Boyatzis, R. E., Rochford, K., & Jack, A. I. (2014). Antagonistic neural networks underlying differentiated leadership roles. *Frontiers in Human Neuroscience, 8*, 1–15.

Bradley, R., & Terry, M. (1952). Rank analysis of incomplete block designs. *Biometrica, 39*, 22–38.

Brendillet, C. N., Tywoniak, S., & Dwivedula, R. (2015). Reconnecting theory and practice in pluralistic contexts: Issues and Aristotelian considerations. *Project Management Journal, 46*(2), 6–20.

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software, 80*(4), 571–583.

Buckingham, A., & Saunders, P. (2004). *The survey methods workbook*. Cambridge, UK: Polity Press.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Science, 1124*(1), 1–38.

Budzier, A., & Flyvberg, B. (2013). *Double whammy—How ICT projects are fooled by randomness and screwed by political intent* (Working Paper). Retrieved from http://arxiv.org/ftp/arxiv/papers/1304/1304.4590.pdf

Burgman, M., Carr, A., Godden, L., Gregory, R., McBride, M., Flander, L., & Maguire, L. (2011). Redefining expertise and improving ecological judgment. *Conservation Letters, 4*(2), 81–87.

Burgman, M. A., McBride, M., Aston, R., Speirs-Bridge, A., Flander, L., Wintle, B., & Twardy, C. (2011). Expert status and performance. *PLoS One, 6*(7).

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making, 5*(3), 182.

Carr, A. J. (2004). Why do we need community science? *Social Natural Resources, 17*, 1–9.

Catenacci, M., Bosetti, V., Fiorese, G., & Verdolini, E. (2015). Expert judgment elicitation protocols. In V. Bosetti & M. Catenacci (Eds.), *Innovation under uncertainty: The future of carbon-free energy technology* (pp. 1–8). Cheltenham, UK: Edward Elgar.

Catenacci, M., Verdolini, E., Bosetti, V., & Fiorese, G. (2013). Going electric: Expert survey on the future of battery technologies for electric vehicles. *Energy Policy, 61*, 403–413.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenber (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 309–336). New York, NY: Routledge.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559–583.

Clemen, R. T. (2008). Comments on Cooke's classical method. *Reliability Engineering & System Safety, 93*(5), 760–765.

Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis, 19*(2), 187–203.

Clemen, R., & Reily, T. (2001). *Making hard decisions with decision tools.* Pacific Grove, CA: Duxbury Press.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155.

Collaros, P. A., & Anderson, L. R. (1969). Effect of perceived expertness upon creativity of members of brainstorming groups. *Journal of Applied Psychology, 53*, 159–163.

COLLEGEdata.com. (2015, June 15). Admissions information & college data. *CollegeData*. Retrieved from http://www.collegedata.com/cs/data/college/college_pg02_tmpl.jhtml?schoolId=xyz

Collins, H. M., & Evans, R. (2007). *Rethinking expertise.* Chicago, IL: University of Chicago Press.

Comer, M. K., Seaver, D. A., Stillwell, W. G., & Gaddy, C. D. (1984). *Generating human reliability errors using expert judgement.* In NUREG/CR-3688. US Nuclear Regulatory Commission Washington, DC.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or Internet-based surveys. *Educational and Psychological Measurement, 60*(6), 821–836.

Cooke, R. M. (1999). *Experts in uncertainty: Opinion and subjective probability in science.* New York, NY: Oxford University Press.

Cooke, R. M. (2015). The aggregation of expert judgment: Do good things come to those who weight? *Risk Analysis, 35*(1), 12–15.

Cooke, R. M., & Goossens, L. L. (2008). TU Delft expert judgment data base. *Reliability Engineering & System Safety, 93*(5), 657–674.

Cooke, R., & Goossens, L. (2000). *Procedures guide for structured expert judgment.* Brussels, Belgium: European Union Atomic Energy Community.

Cornelissen, J. P. (2005). Beyond compare: Metaphor in organization theory. *Academy of Management Review, 30*, 751–764.

Couper, M. P. (2000). Review: Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 464–494.

Crawford-Brown, D. (2001). Scientific methods of human health risk analysis in legal and policy decisions. *Human Health Risk Analysis, 64*(4), 63–81.

Creswell, J. W. (2013). *Qualitative, quantitative, and mixed method approaches.* Thousand Oaks, CA: Sage.

Crosetto, P., & Filippin, A. (2013). The “bomb” risk elicitation task. *Journal of Risk & Uncertainty, 47*, 31–65.

Curtis, A., & Wood, R. (2004). Optimal elicitation of probabilistic information from experts. *Geological Society of London Special Publication, 239*, 127–145.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method. *Management Science, 9*, 458–467.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York, NY: Free Press.

DeGroot, M. H., & Fienberg, S. E. (1982). Assessing probability assessors: Calibration and refinement. In S. S. Gupta & J. O. Berger (Eds.), *Statistical decision theory and related topics III* (Vol. 1, pp. 291–314). New York, NY: Academic.

Delbecq, A. L., & Van de Ven, A. H. (1975). *Group techniques for program planning*. Glenview, IL: Scott-Foresman.

Dennis, A. R., Minas, R. K., & Bhagwatwar, A. P. (2013). Sparking creativity: Improving electronic brainstorming with individual cognitive priming. *Journal of Management Information Systems, 29*(4), 195–216.

Deuja, A., Kohn, N. W., Paulus, P. B., & Korde, R. M. (2014). Taking a broad perspective before brainstorming. *Group Dynamics: Theory, Research, and Practice, 18*(3), 222–231.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology, 53*(3), 497–509.

Dwivedi, Y. K., Ravichandran, K., Williams, M. D., Miller, S., Lal, B., Antony, G. V., & Kartik, M. (2013). IS/IT project failures: A review of extant literature for deriving a taxonomy of failure factors. In Y. K. Dwivedi, H. Z. Henriksen, D. Wastell, & R. De' (Eds.), *Grand successes and failures in IT, public and private sectors* (pp. 73–88). London, UK: Springer.

Dyer, J. S. (1990). Remarks on the analytic hierarchy process. *Management Science, 36*(3), 249–258.

Dzindolet, M. T., Paulus, P. B., & Glazer, C. (2012). Brainstorming in virtual teams. In C. Nunes Silva (Ed.), *Online research in urban and planning studies: Design and outcome* (pp. 138–156). Lisbon, Portugal: IGI Global.

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior, 23*(4), 281–295.

Evans, R. (2008). The sociology of expertise: The distribution of social fluency. *Social Compass, 2*, 281–298.

Fasolo, B., & Bana e Costa, C. A. (2014). Tailoring value elicitation to decision makers' numeracy and fluency: Expressing value judgments in numbers or words. *Omega, 44*, 83–90.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, UK: Cambridge University Press.

Flyvberg, B. (2006). From Nobel Prize to project management: Getting risks right. *Project Management Journal, 37*(3), 5–15.

Flyvberg, B., Holm, M. K., & Buhl, S. L. (2002). Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association, 68*(3), 279–295.

Flyvberg, B., Holm, M. K., & Buhl, S. L. (2005). How (in)accurate are demand forecasts in public works projects? The case of transportation. *Journal of the American Planning Association, 71*(2), 131–146.

Fong, T. G., Gleason, L. J., Wong, B., Habtemariam, D., Jones, R. N., Schmitt, E. M., . . . Inouye, S. K. (2015). Cognitive and physical demands of activities of daily living in older adults: Validation of expert panel ratings. *PM&R*, In Press.

Fox, M. D., Corbetta, M., Snyder, A. Z., & Vincent, J. L. (2006). Spontaneous neuronal activity distinguished human dorsal and ventral attention systems. *Proceedings of the National Academy of Science, 103*(36), 10046–10051.

Frederick, S. "Cognitive reflection and decision making." *The Journal of Economic Perspectives 19, no. 4* (2005): 25–42.

French, S. (1985). Group consensus probability distributions: A critical survey. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. Smith (Eds.), *Bayesian statistics 2*. North-Holland, Netherlands: Elsevierience Publishers.

French, S. (2011). Aggregating expert judgments. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas, 105*(1), 181–206.

Furnham, A. (2000). The brainstorming myth. *Business Strategy Review, 11*(4), 21–28.

Gadzella, B. M., Hogan, L., Masten, W., Stacks, J., Stephens, R., & Zascavage, V. (2006). Reliability and validity of the Watson-Glaser Critical Thinking Appraisal for different academic groups. *Journal of Instructional Psychology, 33*(2), 141–149.

Garcia, C. B., Garcia-Perez, J., & Sanchez-Granero, M. A. (2012). An alternative for robust estimation in project management. *European Journal of Operational Research, 220*(2), 443–451.

Genest, C., & McConway, K. J. (1990). Allocating the weights in the linear opinion pool. *Journal of Forecasting, 9*, 53–73.

Genest, C., & Zidek, J. V. (1986). Combining probability distributions: A critique and annotated bibliography. *Statistical Science,* 114–148.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment.* Cambridge, UK: Cambridge University Press.

Gobble, M. A. (2014). The persistence of brainstorming. *Research Technology Management, 57*(1), 64–70.

Golenko-Ginzburg, D. (1989). PERT assumptions revisited. *Omega, 17,* 393–396.

Grigore, B., Peters, J., Hyde, C., & Stein, K. (2013). Models to elicit probability distributions from experts: A systematic review of reported practice in health technology assessment. *Pharmaco Economics, 31,* 991–1003.

Grinblatt, M., & Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *The Journal of Finance, 64*(2), 549–578.

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology* (Vol. 561). New York, NY: Wiley & Sons.

Groves, R. M., Fowler F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2013). *Survey methodology.* New York, NY: Wiley & Sons.

Guilford, J. P. (1967). *The nature of human intelligence.* New York, NY: McGraw-Hill.

Guilford, J. P., & Guilford, J. S. (1980). *Christensen-Guilford fluency tests: Manual of instruction and interpretations.* Palo Alto, CA: Mind Garden.

Gustafson, D., Shulka, R., Delbecq, A., & Walster, G. (1973). A comparitive study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. *Organizational Behavior and Human Performance, 9*, 280–291.

Hahn, E. D. (2008). Mixture densities for project management activity times: A robust approach to PERT. *European Journal of Operational Research, 188*, 450–459.

Haran, U., Moore, D., & Morewedge, C. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making, 5*(7), 467–476.

Herrerías, R., García, J., & Cruz, S. (2003). A note on the reasonableness of PERT hypotheses. *Operations Research Letters, 31*, 60–62.

Herrerías-Velasco, J. M., Herrerías-Pleguezuelo, R., & van Dorp, J. R. (2011). Revisiting the PERT mean and variance. *European Journal of Operational Research, 210*, 448–451.

Herzog, S., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectic bootstrapping. *Psychological Science, 20*(2), 231–237.

Higgins, J. M. (1994). *101 creative problem solving techniques.* Winter Park, FL: The New Management Publishing Co.

Hill, G. W. (1982). Groups vs. individual performance: Are N + 1 heads better than one? *Psychological Bulletin, 91*, 517–539.

Hogarth, R. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association, 70*, 271–289.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Science, 101*(46), 16385–16389.

Hora, S. C. (2004). Probability judgments for continuous quantities: Linear combinations and calibration. *Management Science, 50*(5), 597–604.

Hora, S., & von Winterfeldt, D. (1997). Nuclear waste and future societies: A look into the deep future. *Technological Forecasting and Social Change, 56*, 155–170.

Jack, A. I., Dawson, A. J., & Norr, M. (2013). Seeing human: Distinct and overlapping neural signatures. *Neuroimage, 79*, 313–328.

Jacobs, C. D., Oliver, D., & Heracleous, L. (2013). Diagnosing organizational identity beliefs by eliciting complex, multi-modal metaphors. *The Journal of Applied Behavioral Science, 49*(4), 485–507.

Jain, K., Mukherjee, K., Bearden, J. N., & Gaba, A. (2013). Unpacking the future: A nudge toward wider confidence intervals. *Management Science, 59*(9), 1970–1987.

Jannis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes.* Boston, MA: Houghton Mifflin.

Jannis, I. L., & Mann, L. (1977). *Decision making.* New York, NY: Free Press.

Johnson, E. N. (1995). Effects of information order, group assistance, and experience on auditors' sequential belief revision. *Journal of Economic Psychology, 16*(1), 137–160.

Jones, C. (2004). Software project management practices: Failure versus success. *Cross Talk: The Journal of Defense Software Engineering*, 5–9.

Jørgensen, M. (2004). A review of studies on expert estimation of software development effort. *Journal of Systems and Software, 70*(1), 37–60.

Jørgensen, M., Halkjelsvik, T., & Kitchenham, B. (2012). How does project size affect cost estimation error? Statistical artifacts and methodological challenges. *International Journal of Project Management, 30*(7), 839–849.

Jørgensen, M., & Shepperd, M. (2007). A systematic review of software development cost estimation studies. *IEEE Transactions on Software Engineering, 33*(1), 33–53.

Kahneman, D. (2011). *Thinking fast and slow.* New York, NY: Farrar, Straus, & Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York, NY: Cambridge University Press.

Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames.* Cambridge, UK: Cambridge University Press.

Kaplan, S. (1990). "Expert information" vs. "expert judgments": Another approach to eliciting/combining using expert knowledge in PRA. *Reliability Engineering and System Safety, 35*(1), 61–72.

Kara, H. (2015). *Creative research methods in the social sciences: A practical guide.* Bristol, UK: Policy Press.

Keith, D. W. (1996). When is it appropriate to combine expert judgments? A critique and annotated bibliography. *Statistical Science, 19*(2), 187–203.

Kennedy, M. K., & Vargus, B. (2001). Challenges in survey research and their implications for philanthropic studies research. *Nonprofit and Volutary Sector Quarterly, 30*(3), 483–494.

Kent, S. (1964). The words of estimative probability. *Intelligence Studies, 8*, 49–64.

Kerzner, H. (2012). *Project management: A systems approach to planning, scheduling, and controlling* (11th ed.). New York, NY: Wiley.

Kerzner, H. R. (2011). *Using the project management maturity model: Strategic planning for project management.* New York, NY: Wiley & Sons.

Kim, B. C., & Reinschmidt, K. F. (2011). Combination of project cost forecasts in earned value management. *Journal of Construction Engineering and Management, 137*(11), 958–966.

Kleibeuker, S. W., Koolschijn, P. C., Jolles, D. D., de Dreu, C. K., & Crone, E. A. (2013). The neural coding of creative idea generation across adolescence and early adulthood. *Frontiers in Human Neuroscience*, 7.

Kohn, N. W., & Smith, S. M. (2011). Collaborative fixation: Effects of others' ideas on brainstorming. *Applied Cognitive Psychology, 25*(3), 359–371.

Koriat, A., Lichtenstein, S., & Fischoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology and Human Learning Memory*, 107–118.

Koskela, L., & Howell, G. (2002). The underlying theory of project management is obsolete. In Proceedings of the PMI Research and Education Conference (pp. 293–302). Seattle, WA: Project Management Institute.

Kraaijenbrink, J. (2010). *Rigor and relevance under uncertainty: Toward frameworks as theories for practice* (Working Paper). Retrieved from http://kraaijenbrink.com/wp-content

/uploads/2012/06/Rigor-and-relevance-under-uncertainty-Kraaijenbrink-20-12-2010.pdf
Krippendorff, K. (2012). *Content analysis: An introduction to its methodology.* Thousand Oaks, CA: Sage.
Kullback, S. (1959). *Information theory and statistics.* New York, NY: Wiley & Sons.
Kynn, M. (2008). The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 171*(1), 239–264.
Labovitz, S. (1970). The assignment of numbers to rank order categories. *American Sociological Review*, 515–524.
Lamn, H., & Trommsdorff, G. (1973). Group versus individual performance on tasks requiring ideational proficiency (brainstorming). *European Journal of Social Psychology, 3*, 361–387.
Larichev, O. I., & Brown, R. V. (2000). Numerical and verbal decision analysis: Comparison on practical cases. *Journal of Multi-Criteria Decision Analysis, 9*(6), 263–273.
Lawrence, M., Goodwin, P., O'Connor, M., & Onkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting, 22*(3), 493–518.
Li, Y. F., Xie, M., & Goh, T. N. (2009). A study of project selection and feature weighting for analogy based software cost estimation. *Journal of Systems and Software, 82*(2), 241–252.
Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Decision Processes, 20*, 159–183.
Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1981). *Calibration of probabilities: The state of the art to 1980*. Eugene, OR: Decision Research.
Lin, S. W., & Bier, V. M. (2008). A study of expert overconfidence. *Reliability Engineering & System Safety, 93*(5), 711–721.
Linsey, J. S., & Becker, B. (2011). Effectiveness of brainwriting techniques: Comparing nominal groups to real teams. In T. Taura, & Y. Nagai (Eds.), *Design creativity 2010* (pp. 165–171). London, UK: Springer.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*, 37–44.

Liu, L., & Napier, Z. (2010). The accuracy of risk-based cost estimation for water infrastructure projects: preliminary evidence from Australian projects. *Construction Management and Economics, 28*(1), 89–100.

Malcolm, D. G., Roseboom, C. E., Clark, C. E., & Fazar, W. (1959). Application of a technique for research and development program evaluation. *Operations Research, 7*, 646–649.

Mantel, S. J., Meredith, J. R., Shafer, S. M., & Sutton, M. M. (2010). *Project management in practice* (4th ed.). New York, NY: Wiley.

Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science, 22*, 1087–1096.

McDonald, S., Zhao, T., & Edwards, H. M. (2013). Dual verbal elicitation: The complementary use of concurrent and retrospective reporting within a usability test. *International Journal of Human-Computer Interaction, 29*(10), 647–660.

Mendell, M., & Sheridan, T. (1989). Filtering information from human experts. *IEEE Transactions on Systems, Man, and Cybernetics, 36*, 6–16.

Merrick, J., van Dorp, J. R., & Singh, A. (2005). Analysis of correlated expert judgments from extended pairwise comparisons. *Decision Analysis, 2*(1), 17–29.

Meyer, M. A., & Booker, J. M. (2001). *Eliciting and analyzing expert judgment*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Morris, P. A. (1977). Combining expert judgments—A Bayesian approach. *Management Science, 23*, 679–693.

Mosleh, A., & Apostalakis, G. (1982). Models for the use of expert opinion. In R. Walker, & V. Covello, *Low-probability high-consequence risk analysis*. New York, NY: Plenum Press.

Murphy, A. (1972a). Scalar and vector partitions of the probability score: Part I. Two-state situation. *Journal of Applied Meteorology, 11*, 273–282.

Murphy, A. (1972b). Scalar and vector partitions of the probability score: Part II. N-state situation. *Journal of Applied Meteorology, 11*, 1183–1192.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J. R., Garthwaite, P., Jenkinson, D., . . . Rakow, T. (2006). *Uncertain judgements: Eliciting experts' probabilities.* Chichester, UK: Wiley & Sons.

Osborn, A. F. (1957). *Applied imagination.* New York, NY: Scribner.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29*, 261–265.

Otto, K., & Wood, K. (2001). *Product design: Techniques in reverse engineering and new product development.* Upper Saddle River, NJ: Prentice Hall.

Parke, B., Hunter, K. F., Strain, L. A., Marck, P. B., Waugh, E. H., & McClelland, A. J. (2013). Facilitators and barriers to safe emergency department transitions for community dwelling older people with dementia and their caregivers: A social ecological study. *International Journal of Nursing Studies, 50*, 1206–1218.

Pascarella, E. T. (1989). *The development of critical thinking: Does college make a difference?* Journal of College Student Development.

Peters, E., Vastfjall, D., Slovic, P., Mertz, C., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*, 407–413.

Plous, S. (1993). *The psychology of judgment and decision making.* New York, NY: McGraw-Hill.

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: Can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management, 29*(2), 245–256.

Premachandra, I. M. (2001). An approximation of the activity duration distribution in PERT. *Computers & Operations Research, 28*(5), 443–452.

Project Management Institute. (2012). *PMI lexicon of project management terms.* Newtown Square, PA: Author.

Project Management Institute. (2013). *A guide to the project management body of knowledge (PMBOK® guide)* – Fifth edition. Newtown Square, PA: Author.

Project Management Institute. (2014, December 1). Hong Kong natural gas pipeline. Retrieved from http://www.pmi.org/Business-Solutions/~/media/PDF/Case%20Study/HK_Pipeline_casestudy_v3.ashx

Project Management Institute. (2015, April). PMI fact file. *PMI Today*, p. 4.

Pulkkinen, U. (1993). Method for combination of expert judgments. *Reliability Engineering and System Safety, 40*, 111–118.

Rauhut, H., & Lorenz, J. (2010). The wisdom of crowds in one mind: How individuals can simulate the knowledge of diverse societies to reach better decisions. *Journal of Mathematical Psychology, 55*(2), 191–197.

Roelofs, V. J., & Roelofs, W. (2013). Using probability boxes to model elicited information: A case study. *Risk Analysis, 33*(9), 1650–1660.

Saaty, T. (1980). *The analytic hierarchy process.* New York, NY: McGraw-Hill.

Sackman, H. (1975). *Delphi critique, expert opinion, forecasting and group processes.* Lexington, MA: Lexington Books.

Sandberg, J., & Tsoukas, H. (2011). Grasping the logic of practice: Theorizing through practical rationality. *Academy of Management Review, 36*(2), 338–360.

Sasieni, M. W. (1986). Note—A Note on Pert Times. *Management Science, 32*(12), 1652–1653.

Shah, J. J. (1998). Experimental investigation of progressive idea generation techniques in engineering design. *Proceedings of the Design Engineering Technical Conference* (pp. 13–16). Atlanta, GA: ASME.

Shah, J. J., Kulkarni, S. V., & Vargas-Hernández, N. (2000). Evaluation of idea generation methods for conceptual design: Effectiveness metrics and design of experiments. *Transactions of the ASME Journal of Mechanical Design, 122*, 377–384.

Shah, J. J., Smith, N. N., & Vargas-Hernández, N. (2003). Metrics for measuring ideation effectiveness. *Design Studies, 24*, 111–134.

Shah, J. J., Vargas-Hernández, N., Summers, J. S., & Kulkarni, S. (2001). Collaborative sketching (C-Sketch)—An idea generation technique for engineering design. *Journal of Creative Behavior, 35*, 168–198.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behaviour and Human Decision Processes, 53*, 252–266.

Simola, K., Mengolini, A., Bolado-Lavin, R., & Gandossi, L. (2005). *Training material for formal expert judgment* (Technical Report 21770). Brussels, Belgium: EU.

Smerek, R. (2014). Why people think deeply: Meta-cognitive cues, task characteristics and thinking dispositions. *Handbook of Research Methods on Intuition*, 3–14.

Söderlund, J. (2004). Building theories of project management: Past research, questions for the future. *International Journal of Project Management 22*(3), 183–191.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 299.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 780.

Soll, J. B., Milkman, K. L., & Payne, J. W. (2014). A user's guide to debiasing. In K. Gideon & G. Wu (Eds.), *Handbook of judgment and decision making* (pp. 924–951). New York, NY: Wiley-Blackwell.

Song, L. (2010). *Earned value management: A global and cross-industry perspective on current EVM practice.* Newtown Square, PA: Project Management Institute.

Specter, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9*(2), 221–232.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis, 30*(3), 512–523.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary.* New York, NY: Oxford University Press.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and brain sciences, 23*(05), 701–717.

Stevens, S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677–680.

Stone, M. (1961). The opinion pool. *Annals of Mathematical Statistics, 32*, 1339–1342.

Subramaniam, K., Kounios, J., Parrish, T. B., & Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive effect. *Journal of Cognitive Neuroscience, 21*(3), 415–432.

Sudhakar, G. P. (2013). The Key functions and best practices of software product management. *Sprouts: Working Papers on Information Systems, 13*(2), 13–21.

Susel, I. (2011). *Issues to be addressed when eliciting expert judgments for operational risk scenarios.* Washington, DC: Department of Homeland Security.

Svejvig, P., & Andersen, P. (2015). Rethinking project management: A structured literature review with a critical look at the brave new world. *International Journal of Project Management, 33*(2), 278–290.

Szwed, P. S. (2014). Making the case for (re)defining expert judgment (elicitation) in project management. *Proceedings of the PMI Research and Education Conference.* Portland, OR: Project Management Institute.

Szwed, P. S., van Dorp, J. R., Merrick, J. R., & Singh, A. (2006). A Bayesian paired comparison approach for relative accident probability assessment with covariate information. *International Journal of Operational Research, 169*(1), 157–177.

Takeuchi, H., Taki, Y., Hashizume, H., Sassa, Y., Nagase, T., Nouchi, R., & Kawashima, R. (2011). Failing to deactivate: The association between brain activity during working memory task and creativity. *Neuroimage, 55*(2), 681–687.

Taylor, D. W., Berry, P. C., & Block, C. H. (1958). Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administrative Science Quarterly, 3*, 23–47.

Teigen, K. H., & Jørgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology, 19*(4), 455–475.

Tetlock, P. E. (2005). *Expert political judgment.* Princeton, NJ: Princeton University Press.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*(4), 273–281.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275–1289.

Torgersen, W. S. (1958). *Theory and methods of scaling.* New York, NY: Wiley.

Trendowicz, A., Munch, J., & Jeffery, R. (2011). State of practice in software effort estimation: A survey and literature review. In *Software Engineering Techniques* (pp. 232–245). Berlin, Germany: Springer.

Tsai, C. I., Klayman, J., & Hastie, R. (2008). Effects of amount of information on judgment accuracy and confidence. *Organizational Behavior and Human Decision Processes, 107*(2), 97–105.

Tumonis, V., Šavelskis, M., & Žalytė, I. (2013). Judicial Decision-Making From An Empirical Perspective. *Baltic Journal of Law & Politics, 6*(1), 140–162.

Tversky, A., & Kahneman, D. (1974). Judgments under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.

Uddin, L. Q., Kelly, A. M., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2009). Functional connectivity of default mode network components: Correlation, anticorrelation, and causality. *Human Brain Mapping, 30*(2), 625–637.

U.S. Environmental Protection Agency. (2009). *Expert Elicitation Task Force White Paper*, GPO: Washington, DC.

Van de Ven, A. H., & Johnson, P. E. (2006). Knowledge for theory and practice. *Academy of Management Review, 31*(4), 802–821.

van Dorp, J. R. (2012). Indirect parameter elicitation procedures for some distributions with bounded support–with applications in Program Evaluation and Review Technique (PERT). *Structure and Infrastructure Engineering, 8*(4), 393–401.

VanGundy, A. B. (1988). *Techniques for structured problem solving.* New York, NY: Can Nostrand Reinhold Co.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science, Short Report, 19*, 645–647.

Wachowicz, T., & Blaszczyk, P. (2013). TOPSIS based approach to scoring negotiating offers in negotiation support systems. *Group Decision & Negotiation, 22*(6), 1021–1050.

Walker, K. D., Evans, J. S., & MacIntosh, D. (2001). Use of expert judgment in exposure assessment. *Journal of Exposure Analysis and Environmental Epidemiology—Part I, 11*, 308–322.

Watson, G. (1980). *Watson-Glaser critical thinking appraisal.* San Antonio, TX: Psychological Corporation.

Welsh, M., Lee, M., & Begg, S. (2008). More-or-less elicitation (MOLE): Testing a heuristic elicitation method. *Proceedings of the 30th Annual Meeting of Cognitive Science Society*, 493–498.

Welsh, M., Lee, M., & Begg, S. (2009). Repeated judgments in elicitation tasks: Efficacy of MOLE method. *Proceedings of the 31st Annual Meeting of Cognitive Science Society*, 1529–1534.

Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: how to reduce overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1167.

Woloshin, S., Schwartz, L. M., Moncour, M., Gabriel, S., & Tosteson, A. N. (2001). Assessing values for health: Numeracy matters. *Medical Decision Making, 21*, 382–390.

Woods, L. E., & Ford, J. M. (1993). Structuring interviews with experts during knowledge elicitation. In K. M. Ford & J. M. Bradshaw (Eds.), *Knowledge acquisition as modeling* (pp. 71–90). New York, NY: Wiley.

Yates, J. F. (1994a). Analyzing the accuracy of probability judgments for multiple events: An extension of the covariance decomposition. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 381–410). Chichester, UK: Wiley.

Yates, J. F. (1994b). Subjective probability accuracy analysis. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 381–410). Chichester, UK: Wiley.

Yin, R. K. (2003). *Case study research: Design and methods* (Vol. 5). Thousand Oaks, CA: Sage.

Zikmund-Fisher, B. J., Smith, D. M., Ubel, P. A., & Fagerlin, A. (2007). Validation of the subjective numeracy scale (SNS): Effects of low numeracy on comprehension of risk communication and utility evaluations. *Medical Decision Making, 27*, 663–671.

Expert Judgment IN Project Management

Narrowing the Theory-Practice Gap

Expert judgment is a major source of information that can provide vital input to project managers, who must ensure that projects are completed successfully, on time, and on budget.

Too often, however, companies lack detailed processes for finding and consulting with experts—making it hard to match the required know-how with the project at hand. In *Expert Judgment in Project Management: Narrowing the Theory-Practice Gap*, Paul S. Szwed provides research that will help project managers become more adept at using expert judgment effectively.

The author explores the use of expertise in several sectors, including engineering, environmental management, medicine, political science, and space exploration. He then looks at the informal state of expert judgment and its underutilization in the management of projects.

Szwed's critical recommendations can help project managers improve the way they select, train, and work with experts to increase the odds of any project's success.

PMI®
Project Management Institute

ISBN: 978-1-62825-116-6 U.S. $24.95
52495
9 781628 251166