1-1-2014

The Road Taken: A Between And Within Subjects Study Of Intervention Selection Decisions By Performance Improvement Professionals

Hillary Leigh
Wayne State University,

THE ROAD TAKEN:
A BETWEEN AND WITHIN SUBJECTS STUDY OF INTERVENTION SELECTION DECISIONS BY PERFORMANCE IMPROVEMENT PROFESSIONALS

by

HILLARY N. LEIGH

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2014

MAJOR: INSTRUCTIONAL TECHNOLOGY

Approved by:

_____

Advisor Date

_____

_____

_____

_____

The Road Not Taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.
—Robert Frost—

DEDICATION

For my parents, Wayne and Carol Andrei.

Without their continual support and encouragement,
I would not have had the courage to begin this journey
nor the wherewithal to finish it.

ACKNOWLEDGEMENTS

I am grateful for the guidance and support of my committee chair, Dr. Ingrid Guerra-López, who patiently afforded me independence and trust during this process. In this work and my life, I am grateful for her mentorship and friendship. Through her, I learned the importance of organizational work in society and a better way for accomplishing the results that are demanded by it. I would also like to thank the other Instructional Technology faculty members on my committee, Dr. James L. Moseley and Dr. Timothy W. Spannaus, for their knowledge, effort, and guidance. Their thoughtful questions helped to greatly refine this work. Last, but not least, I am obliged to my cognate adviser, Dr. Susan Vineberg. Without her introduction to subjective probability and Bayes' theorem, this study would have been far less rigorous and I wouldn't spend nearly so much time deciding which wine to bring to a dinner party.

I am very thankful to Dr. Samiran Ghosh, who consulted on both the frequentist and Bayesian analyses. This work and my understanding of the issues in conducting multiple, varied statistical approaches are better for his expert consultancy. Special thanks are due to Peter Jung who supported me during various phases of the study, most especially in coding of the pilot data. Thanks are also due to Dr. Bulent Ozkban who advised on frequentist tests early on; Michele Norris, who often helped me ferry paperwork from more than 2,000 miles away; and to Nicole Cowell, Laura Berman, Michelle Bondurant, Stephen Bogdanov, Michael Stoutsenberger, Jeanne Dang, Jennifer Ralph and Winnie Shu for their input during my final preparations for defense.

I would like to acknowledge April Davis, Executive Director, and Dr. Gay Bruhn at ISPI for allowing access to CPTs as the sample frame and for their assistance in inviting study participation. The longevity of a professional organization is dependent upon its support of reflective, empirical inquiry; ISPI's continued willingness to support Doctoral research embodies this characteristic.

I am especially grateful for the motley crew of colleague-friends who supported me during this process. Drs. Debra M. Smith, Stacey L. Schepens, Chong Y. Kim, Kara Y. Snow, and Simon Funge were a thoughtful sounding board throughout, providing advice, encouragement, and a sympathetic ear when required. I learned by example and their work inspired me to do my best.

I would especially like to thank Dr. Doug Leigh for his support through the darkest periods of this journey. He treated each epiphany as if it were both my first and last. When I was tangled up about how to proceed, he would listen for a few minutes then offer a simple and elegant solution, seemingly without effort. Days or weeks later, one would find me knee deep in the literature, only to conclude that the suggestion was the best option. The dissertation process would have been a lot shorter had I taken his advice more readily. Either way, I am deeply indebted to him.

Finally, I am appreciative to the subjects who participated in my study. Without them, I would have learned nothing.

PREFACE

Although this is an introductory page, it is one of the last to be written. As such, it is profoundly personal and sweet to apply the letters to the page. In many ways, I am surprised that this is the case.

This work, and the Doctoral degree that it culminates, began as an intellectual journey. When I looked around at those who were doing work I envied, almost all carried the letters *Ph.D.* on their business cards. Therefore, as I embarked, I viewed a Doctoral degree as a means to an end of obtaining more gratifying, meaningful, and thoughtful work in my career. As I advanced in my studies, I was perplexed by the lack of empirical study on intervention selection and was frankly enamored with the promise of a decision theoretic for framing research in that area.

As I soon learned, this meant revisiting long forgotten (on my part) conventions of symbolic logic and mathematical notation as well as reading works that were written more than 250 years ago. Paths, worn or otherwise, were hard to come by and it was in the weeds of this complexity that the more personal aspects of my journey began.

Writing came less easily. There were days when completing a paragraph was considered a “monumental” “success” “!”. I would sometimes awaken in the middle of the night with an epiphany and re-write entire sections, only to abandon most of the revisions in the light of the next day. In this dark wood, I somehow navigated to a defensible proposal of a study and the approval of my committee to collect data ushered in new hope. Alas, it was short-lived.

Perfectionism had egged me into several phases of highly rigorous instrument validation techniques. Unfortunately, the real world did not always comply with gusto to my plans: there were delayed approvals, last minute changes, modifications, misplaced paperwork, locked doors, disoriented deliverymen, complex analyses, and the rest of my life to slow me down. One day I caught myself pontificating aloud about what I would do *"if* I ever finished my dissertation." I probably clapped my hand over my mouth at the time, but the words had already been said many times over in my mind: I had long wondered whether I would find answers to my research questions and had genuinely begun to doubt whether having the answers mattered at all.

As it turned out, *the answers made all the difference*. The statistical results reignited my curiosity and desire to know. Instead of wondering '*if* I would finish', I began to wonder all sorts of other *if's* instead. Words, sentences, sections—even chapters—came more readily. As I was nearing completion of Chapters 4 and 5, confidence replaced self-doubt, clarity settled in where uncertainty had set up semi-permanent residence, and the subtle suggestion of insight began to glitter around the edges.

And in the end—I learned it wasn't about the end at all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

In the classical view, the arts and sciences are diametrically opposed to one another. According to Richmond (1984), the traditional perspective distinguishes science as the rational pursuit of generalized knowledge and truth. Scientific work includes objective and precise methods, along with attention to the detail required for replication. Scientists themselves ought to remain impartial and emotionally detached from the issues they study. In a way, art is characterized as the "anti-science." It is inextricably tied to the artist, and is valuable for its own sake or beauty. The creation of art is characterized as an emotional process where the artist is passionately, even haphazardly, engaged with their work—seeking the realization of their vision or inspiration (pp. 81-82).

Although the strict dichotomy between art and science may be questioned, as fields professionalize they tend to describe their practice along this continuum. Performance improvement (PI) is a "systemic and systematic process for assessing and analyzing performance gaps, planning improvements in performance, designing and developing efficient, effective and ethically justifiable interventions to close performance gaps, implementing the interventions, and evaluating all levels of results" (Guerra, 2001b, pp. 10-11). In keeping with the development of other professions, the ideal nature of PI practice as a science or an art has elicited a contentious debate (Leigh, 2004; Shrock & Coscarelli, 1981).

In the midst of this controversy, only a few authors have advocated artistic approaches. A conservative position in this vein, Westgaard (1997) asserted that practitioners ought to be heurists by thinking holistically about solving performance

problems. Lewis (2005), on the other hand, put forth more zealous support for an artistic paradigm by suggesting that organizations ought to shift away altogether from a performance-orientation to an artistic focus on social interaction and the meaning of work. Some authors, including Robinson and Robinson (2006), and Van Tiem, Moseley, and Dessinger (2000), and Van Tiem, Moseley, and Dessinger, (2001) describe PI as *both* an art and a science. In fact, Robinson and Robinson (2006) contend that it is only by acknowledging both the artistic and scientific elements that practitioners can engage clients and achieve organizational results. The most prevalent view is that the practice of performance improvement ought to be scientific in nature (Clark & Estes, 2000; P. Dean, 1997; Farrington & Clark, 2000; Foshay, 2000; Gilbert, 1992; Gilbert & Gilbert, 1989; Kaufman & Clark, 1999; Mitchell, 1993; Munn, 2005; Romme & Damen, 2007).

Problem Statement

While the view that performance improvement ought to be scientific is the predominant perspective, review of the published literature—as an indicator of professional practice—suggests a lack of emphasis on empirical data presented in:

1. the then National Society for Performance Improvement's (NSPI) publications (Lindsley, as cited in Binder, 1995);
2. Stolovitch and Keeps' (1992) edition of the *Handbook of Performance Technology (*Binder, 1995*)*; and
3. the articles presented in Performance Improvement Quarterly (PIQ) between 1997 and 2000 (Klein, 2002).

Furthermore, a multitude of gaps exist between research and practice. For example, some practitioners hold a number of beliefs that when examined systematically, are not supported by findings from research (Farrington & Clark, 2000; Hollenbeck, DeRue, &

Guzzo, 2004; Stolovitch, 2000). Additionally, practice appears to be lagging behind research: Colbert, Rynes, and Brown (2005) studied human resource managers' use of informational sources and agreement with empirical research findings. Here, the only source of information that was significantly related with agreeing with research findings was academic reading (i.e. journals); however only 2% of the subjects reported regularly reading one of the journals listed and 75% reported never reading any of them (p. 319).

Still, more recent developments seem to suggest that a commitment to a scientific approach is increasing, as the field's literature-base is growing in both quantity and quality. Stolovitch and Keeps (2006) reported a tenfold increase in the general number of PI publications (i.e. books, periodicals, and articles) since 1992. Guerra-López and Leigh (2009) performed a content analysis of articles published between 1997 and 2006 in the field's primary journals, PIQ and Performance Improvement (PIJ), and found increased focus on rigorous performance measurement in both journals. During that time, 34% of the articles in PIJ related to performance measurement while PIQ showed an even higher rate, as 186 of 247 articles (75%) focused on performance measurement, inclusive of evaluation articles (p.103). Contemporaneously, the percentage of the field's published research that was empirical increased 36% between 1997 and 2000 (Klein, 2002) and 54% between 2001 and 2005 (Marker, Huglin, & Johnson, 2006).

Alongside this growth, arguments for a transition to PI as a scientific endeavor continue. Perhaps the most voracious and notable argument in favor of this position comes from a series of works by Clark and Estes (1998, 2000, 2003). Within this argument, craft and technology are distinguished by their overall nature, purpose, ability

for disconfirmation, and in the isolation of factors that are involved in performance. An especially stark contrast is made between the sources of knowledge typically employed in these approaches. Specifically, Clark and Estes contend that rather than drawing on sources of knowledge generally used in technical approaches such as principles, theoretical models, experimentation, and clear specification of problems, when practitioners select interventions, they tend to rely upon sources often used in craft-based approaches—such as luck, personal expertise, insight, the experiences of others, and trial-and-error (Clark, 2003; Clark & Estes, 1998; Clark & Estes, 2000; Clark & Estes, 2002; Kaufman & Clark, 1999).

Ironically, these claims about practitioner beliefs and their usage of various sources of knowledge are supported almost exclusively through anecdotes. In fact, a review of the field's primary serial publications and journals dating back to 1962 (i.e. PIQ, PIJ, Performance and Instruction, Improving Human Performance Quarterly, Improving Human Performance, NSPI Journal, and the NSPI Newsletter) revealed no studies directly examining this issue. This finding is empirically supported by Huglin, Johnson, and Marker's (2007) three-round Delphi study designed to gain consensus on a research agenda in the field, which found this as one area of import. Specifically, the research question identified was, "What sources of evidence (other than analysis of the client situation at hand) do practitioners draw upon (e.g. experience, research literature, anecdotal case study reports, collegial consultation, etc.) when formulating diagnoses and client intervention plans?" (p. 87).

An important construct within this question is the concept of 'evidence'. Traditional discussions of evidence-based practice roughly equate the term with findings from research (Rycroft-Malone et al, 2004). However, when one employs a broader notion of 'evidence' as all that "seems to bear upon the truth of a proposition or the likelihood of a particular outcome" (based on Conee, 2004), review of the literature yields empirical research that deals with the issues of the nature of practice and the usage of various sources of evidence more implicitly.

A notable example is a content analysis performed by Guerra (2001a), that revealed that only 21% of the articles in PIJ between January 1998 and June 1999 were scientifically-based in systematic analysis and research. Most of the relevant studies of professional practice seek to answer questions about what activities instructional designers actually perform in practice, how frequently they do so, and which underlying principles they value. Of course, one limitation of studies of this type is that while they focus on the general activities performed, they do not provide the level of detail required to understand the decision-making of successful designers (Christensen & Osguthorpe, 2004). Instead, they often attempt to derive implicit heuristics or underlying assumptions from designer practice (e.g. Kirschner, van Merrienboer, Sloep, & Carr, 2002; Visscher-Voerman & Gustafson, 2004). Of the few studies of performance improvement generalists that are relevant, Harless (1995) addresses the issue concerning professional preparation, while Korth (1997, 2000) emphasized internal sources of information and the creative nature of human resource development planning processes.

The study was formulated as a partial response to these problems.

Purpose and Research Questions

Thus, a broad aim of this mixed design (within-and-between-subjects) repeated measures study is to better understand performance improvement professionals' intervention selection decisions. In particular, the study attends to beliefs related to the likely success of an intervention for resolving a performance discrepancy and changes in these beliefs as evidence is received. Additionally, the role of self-reported familiarity with interventions is examined, as is the propensity for professionals to change their mind about which intervention they prefer during this process.

A dynamic, web-delivered questionnaire instrument was used to address these aims. As an introduction to the study's main research questions, Figure 1 provides an overview of the general sequence of core events in the questionnaire:

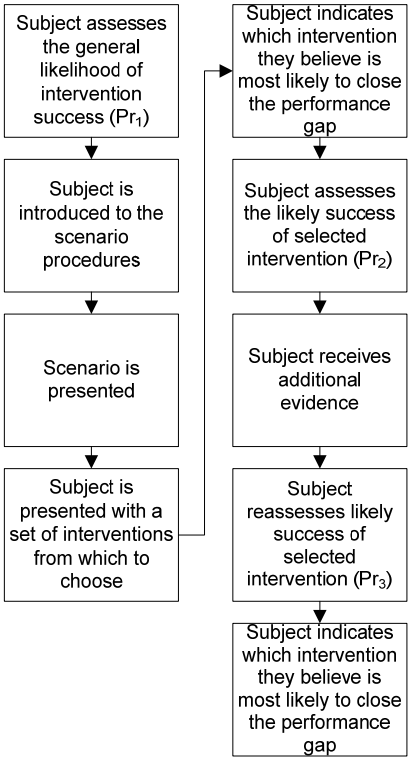


Figure 1. Sequence of questionnaire events.

A more detailed sequence and other methodological issues related to the study's design, population, sampling, administrative procedures, and statistical analysis, will be addressed in the Methods chapter. This sequence has been designed to facilitate study of the following research questions:

1. As they receive evidence, what changes occur in PI professionals' assessments of likely intervention success?
2. Which types of evidence do PI professionals find most persuasive?
3. As they receive evidence, do PI professionals change their minds about what intervention is most likely to succeed?

4. Can PI professionals' changes of mind about what intervention is most likely to succeed be predicted by the nature of evidence received or assessments of likely intervention success?
5. Between instances and non-instances of PI professionals changing their minds about which intervention is most likely to succeed, are there differences in self-reported familiarity with interventions?
6. To what extent (if at all) are PI professionals' assessments of likely intervention success related with self-reported familiarity with interventions?
7. Do periods of practice in probabilistic reasoning influence professionals' assessments of likely intervention success?

Definition of Terms

Agreement. This term identifies the degree to which evidence supports a decision-maker's initially preferred position (based on Chapman, 1973). In this study evidential agreement varied in two ways, the evidence may be (a) supportive of an initial choice or (b) infirming of an initial choice. The latter should be construed as counterevidence that directly points away from the suitability of an initial course of action.

Belief. According to Bandura (1997) and Gagné, Wager, Golas, and Keller (2005) beliefs are individual factors that influence behavior. Schitzgebel (2008) characterizes a belief as an attitude toward a particular proposition such that that proposition is taken to be true; such a representation limits belief to cases of practical certainty. The present

study views beliefs as attitudes toward the truth of a proposition, but these beliefs may come in varying degrees.

Confirmation. There is mixed usage of this term in the literature. Among Bayesian confirmation theorists, 'confirmation' occurs when the probability of a hypothesis is greater after the receipt of some evidence than the probability before it was received (Earman, 1992). But when it is used it is in regard to confirmation bias research, it is generally equated with *agreement*—as defined above.

Decision. A multi-temporal process that begins with the recognition of a problem or choice and ends with the selection of a course of action (based on Mintzberg, Raisinghani, & Theoret, 1976; Nutt, 1984; Witte, 1972).

Evidence. Traditional discussions of evidence-based practice roughly equate 'evidence' with findings from research (Rycroft-Malone et al, 2004). A broader notion is used in this study: as all that "seems to bear upon the truth of a proposition or the likelihood of a particular outcome" (based on Conee, 2004).

Intervention. An individual or set of means used to eliminate or reduce a discrepancy between the achievement of a worthwhile goal and current results of performance (based on Farrington & Clark, 2000).

Intervention selection. Generally, intervention selection involves the decision about which individual or set of means ought to be recommended to a client for implementation to resolve a discrepancy in performance. The selection of an optimal solution involves a variety of factors, including relationship with causes for a discrepancy, likelihood of resolving it, as well contextual factors that may affect

acceptability of the intervention to the client. For the purposes of this study, intervention selection is exclusively concerned with the decision about which intervention is most likely to resolve a performance discrepancy.

Performance improvement. "A systemic and systematic process for assessing and analyzing performance gaps, planning improvements in performance, selecting, designing and developing efficient, effective and ethically justifiable interventions to close performance gaps, implementing the interventions, and evaluating all levels of results" (Guerra, 2001b, pp. 10-11).

Prior and posterior. These terms are used to modify 'probability' as a means for communicating the temporal relationship with evidence. Applying Salmon (1967), prior probability precedes the receipt or collection of evidence, while posterior probability follows it. In the literature, prior probability may also be referred to as 'base-rates', 'initial', 'a priori', or 'Pr (h)'. Posterior probability may also be discussed as 'a posteriori' or 'Pr ($h|e$)'. Measuring changes between posterior probabilities and prior probabilities is somewhat problematic, as results are sensitive to the measure selected (Fitelson, 1999). Among the options, Fitelson argues that there is no compelling reason to prefer either the difference measure *(d)* or the log likelihood ratio measure *(l)* (p. S371). However, probability theory suggests the superiority of l because it normalizes differences in interval length for prior and posterior probabilities (Ghosh, personal communication, December 3, 2010). Therefore, this study used log likelihood ratios as a measure of changes between posterior probabilities and prior probabilities.

Probability. As will be discussed in the following chapter, probability is not a mere mathematical equation. Here, it is instead taken as a relationship between varying degrees of belief and rational preference (Hájek, 2007, section 3.5). When discussed as a construct, *probability* is abbreviated as *Pr*. When *p* values are reported, the standard abbreviation (*p*) is used.

Professional. There are several hallmark characteristics of a professional, including professional association, belief in the value of one's work to society, commitment to self-regulation of the field, a sense of calling to the work, and a belief in autonomous decision-making (Goode, 1957; Hall, 1968). Applying a portion of these characteristics, the operational definition of "professional" in this study is limited to certified performance technologists (CPTs). The justification for this decision will be addressed in greater detail in the sections describing the target population and sample, in the third chapter.

Science and craft. According to Clark and Estes (1998, 2000, 2002), "science" involves systematic inquiry into the way the world works; "craft" on the other hand, involves techniques aimed at solving specific problems. In this study both terms are used to characterize types of evidence that are reported by experts as those that are frequently employed in either approach, as shown in Table 1.

Table 1

Description of Scientific and Craft-based Evidence

Nature of Evidence	Description
Scientific	Findings of prescriptive research Case study Graphical representation of cause and effect Subject matter expert consultation Examining results of implementation of intervention designed to meet similar objectives Pilot testing a prototype
Craft-based	Hunch Client interview Brainstorming with a client Internet research Memories of past experience Editorial article

Importance of the Study

On a practical level, there are several reasons that justify the importance of a study to answer these questions, including the increasing complexity of the intervention selection decision, the relationship of intervention recommendations with organizational decisions, and higher societal expectations for organizations to deliver results and substantiate that they do so. Additionally, this study addresses a few issues of professional and theoretical import.

The intervention selection decision became more complex as the profession trended away from emphasis on instruction and training toward a focus on performance in the 1990s (Guerra, 2001a; Guerra, 2001b; C. Hutchison, 1989; Larson & Lockee, 2004; Reiser & Dempsey, 2007; Robinson & Robinson, 1996; Tovar, Gagnon, & Schmid, 1997). Gayeski (1995) found that six of 10 former NSPI presidents surveyed

reported that the field was moving away from a sole emphasis on instructional solutions toward other interventions and a results-orientation; Tovar et al. (1997) noted equal representation in the field between training and other types of interventions. In fact, there was at least a 7% increase from 1994 to 2005 in professionals whose education specialized in performance improvement (Larson, 2005), and as early as 1995, Dean found that despite human resource and training professionals being the primary originators of performance improvement projects, subjects did not use training on a regular basis.

More recently the American Society for Training and Development (2008) stated that those organizations who demonstrate clear linkages between learning and performance report that 41.3% of the resources utilized within the learning function are devoted to non-training solutions, on average. However, a content analysis performed by Jang (2008) still revealed heavy emphasis on instructional interventions in the literature.

In regard to the usage of interventions in the field, practitioners have progressed away from intervention specialization toward generalized expertise with a variety of possible interventions (Hutchison & Stein, 1998; Langdon, 1997b; Mager, 1992). In the midst of its development, Mager (1992) described the trend:

> …[T]he field became even bigger…when practitioners began to realize that instruction isn't the only way to improve human performance, and when they realized that instruction is seldom the remedy of choice. This development, the realization that there are many variables (and their associated specialties) that impact human performance, is leading to the evolution of the generalists, of the

> professional generalists who can assess the larger situation, who can analyze (diagnose) problems and prescribe solutions—solutions that either they themselves or others will then be assigned to execute." (p. 58)

Logically, alongside an increase in the awareness of a larger set of relevant variables and intervention methods comes an increasing number of questions about how best to decide among them (Ormerod, 1997).

Not only has there been recognition in the field that more types of interventions are possible, but also that more interventions are required. Several authors have noted that resolving a performance discrepancy often requires a complex arrangement of *multiple* interventions, including Harless in Langdon (1997b); Hutchison and Stein (1998); Langdon, Whiteside, and McKenna (1999); Medsker, Hunter, Stepich, Rowland, and Basnet, (1995); and Watkins (2007b). Shifting away from a single solution perspective complicates intervention selection because it may require making trade-offs between particular goals (Quinn & McGrath, 1982). Furthermore, the practitioner must consider how several interventions may interact with one another (Herem, 1979; Langdon et al., 1999).

Another reason that supports the importance of this study is that intervention selection decisions result in recommendations to organizational decision-makers (Watkins, & Wedman, 2003). On this view, intervention selection has implications for organizational decisions, which are inherently complex due to (a) the desire to achieve multiple goals at once, (b) the challenge of identifying multiple viable alternatives, (c) issues of measuring intangible consequences, (d) the long term effects of an action, (e)

multiple stakeholders, (f) lack of certainty and inherent risk, (g) in some cases, critical risks (e.g., "life and limb"), (h) interdisciplinary nature of decisions, (i) multiple decision makers, (j) tradeoffs, (k) varying attitudes towards taking risk, and (l) the inter-relatedness and sequential nature of decisions (Keeney, 1982).

Alongside this issue, the stakes for these organizational decisions are higher. For quite some time, society has demanded that organizations deliver meaningful results (Kaufman & Watkins, 2000; Kaufman, Watkins, Triner, & Smith, 1998; Kukalis, 2009). In the wake of this and several organizational scandals, there has been an increased focus by individuals and organizations on corporate social responsibility (Lindgreen, Swaen, & Johnston, 2009; Wayne, 2009). Organizations are not only required to deliver results, but also credible evidence that they do so (Donaldson, 2006; Donaldson, Christie, & Mark, 2009). Performance improvement professionals, as facilitators of this process, have an obligation to make use of the best available evidence when making their recommendations (Kaufman & Clark, 1999; Thomas, 2006); without empirical inquiry into the relationship between perceptions and usage of evidence and the intervention selection decision, it is difficult to determine whether or not this is the case.

This point leads to a matter of importance for the performance improvement profession. On an ongoing basis, the field struggles with issues of legitimacy and the potential for continued survival. The nature and severity of these problems are illustrated in the ongoing concern for them expressed in the literature. Swanson (1988) suggested that a decreasing commitment to research in human performance technology was a "life-or-death matter" for the field (as cited in Huglin, Johnsen, and Marker, 2007). This point

is echoed in Kaufman and Clark's (1999) ominous warning that we must "establish our effectiveness or fade" due, in part, to our operation in "fad, fashion, and what others are doing" rather than "scientifically, empirically, or research-based practice" (pp. 13-14). More recently, Pershing, Lee, and Cheng (2008) found that despite rapid growth in the field of Performance Improvement, many organizations still do not utilize its systematic processes for resolving performance discrepancies. This point is especially critical given management's preference for reports of higher-level results to support training decisions (D. D. Chapman, 2004; Kusy, 1988). Another key finding related to this point is due to Mattson (2003), who studied 233 training managers' decisions and found a significant effect between report type (i.e., utility, critical outcomes, or anecdotal evidence) and perceived usefulness.

Returning to the present study's importance, the utility of the findings for supporting (or disconfirming) claims regarding the scientific nature of intervention selection in performance improvement remains unclear, until more empirical data are collected. However, at a minimum, pursuing these questions in the first place provides some indication of support for a basis of the field in empirical study. As a final point on the potential practical importance of this study, only a few examples of systematic approaches to intervention selection have been developed, yet empirically-derived models of intervention selection do not exist (Langdon, 1997, 2003). While this research falls short of establishing a prescriptive model for this process, it attempts a preliminary step in that direction in its examination of current practice.

As discussed in the problem statement section, there are only a few studies of the intervention selection process and even fewer that center on multiple types of interventions. The present study is theoretically novel as it emphasizes intervention selection as a decision-making process, rather than a sequence of lockstep procedures. Moreover, the study's focus on changes in beliefs during intervention selection is unique within the field of performance improvement.

Theoretical Framework

As noted in the introduction, this study is generally underpinned by theories of art, science, and professionalization. Within the process of PI, the study focuses on the intervention selection process, especially the concepts of deferred choice, consideration of options, results-orientation, and use of intervention classes. Additionally, intervention selection—and other activities involved in performance improvement—can be viewed as *decisions* (Converse & Weaver, 2008; Watkins, 2007a). In fact, it has been argued that the field ought to be considered through the lens of decision-making and its theory-base (Chermack, 2003; Holton & Naquin, 2005).

With this in mind, the study draws on a basic form of Bayes' theorem, derived from a similar presentation by Lynch (2007):

$$\Pr(h|e) = \frac{\Pr(e|h) \times \Pr(h)}{\Pr(e)}$$

This equation may be read as "the probability (Pr) of a hypothesis, *h,* given some evidence, *e,* is equal to the probability of the evidence given the hypothesis multiplied by the probability of the hypothesis, divided by the probability of the evidence'. Put very

simply, the posterior probability of a hypothesis is dependent upon the probability of the evidence that we have for it (Dufty, 2007).

Bayes' theorem is also connected with accounts of evidential support and theories of confirmation, in the form of three key concepts: (a) hypotheses (and theories) are confirmed relative to individual decision-makers' degree of belief in their veracity, i.e., confirmational relativity; (b) an individual's degree of belief in a particular hypothesis ought to be a function of the body of evidence relevant to it, i.e. evidential proportionism; and (c) decisions about what action to take or what theories to accept or what hypothesis to hold true--occur over time and our views about how likely it is that a particular approach will succeed changes as we accumulate evidence, i.e. incremental confirmation (Joyce, 2003; Hawthorne, 2005).

Here, "confirmation" is distinguished from "acceptance," the latter being equated with certainty or practical verification (Hempel, 1945a, 1945b). Thus, evidence, *e,* may be said to confirm a hypothesis, *h*, if and only if Pr (*h*|*e*) is greater than Pr (*h*). By extension, the opposite is true. Evidence disconfirms a hypothesis, if and only if Pr (*h*|*e*) is less than Pr (*h*) (Earman, 1992). By extension, posterior judgments of probability are dependent upon both evidence *and* prior probability assessments (Blackburn, 2005).

More accurately, if one is seeking to be rational about the beliefs they hold about a theory, then their posterior judgments *should be* apportioned to both the evidence and one's prior beliefs. In reality, there are a number of empirical problems for updating probabilities on the basis of new evidence (Ouwersloot, Nijkamp, & Rietveld, 1998), and heuristics may result in systematic, measurable errors (Tversky & Kahneman, 1974).

Taken together, these conceptual, theoretical, and empirical bases form the framework for the study's variables and research design, as illustrated in Figure 2.

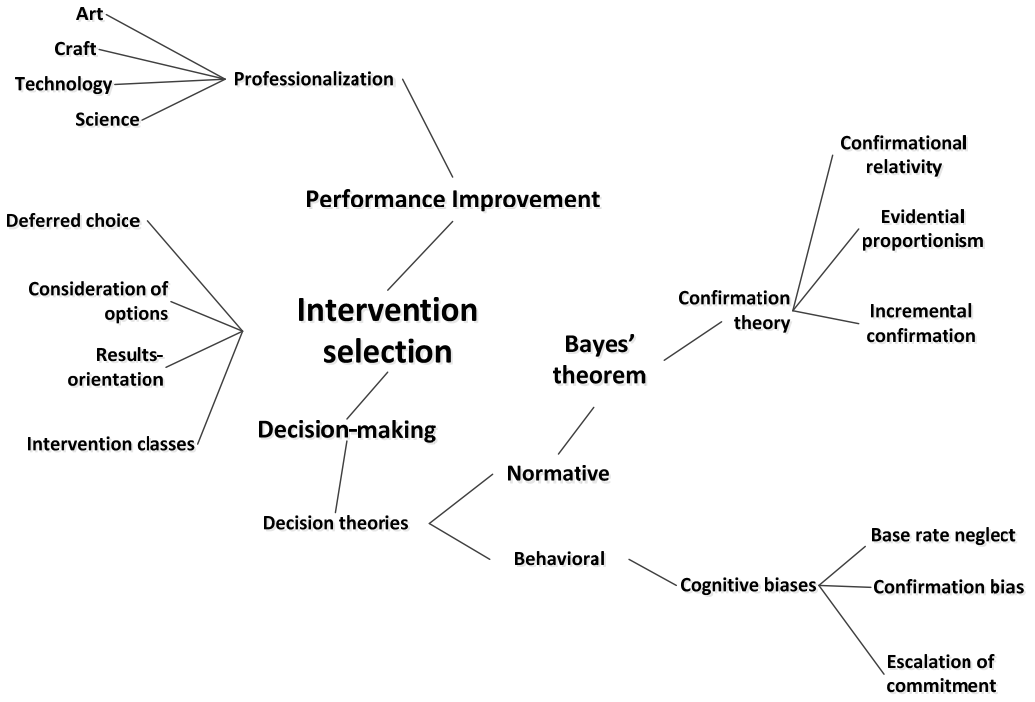


Figure 2. Theoretical framework.

Study Delimitations and Limitations

The practical achievement of this study's aims required delimitation. First, the study focused exclusively on the usage of evidence as it is weighed in making intervention selection decisions. Other aspects of the performance improvement process (e.g., analysis, intervention design, and evaluation) were excluded. Furthermore, broad-based analysis of the interventions selected by PI professionals for the purposes of conclusions about which interventions are selected was outside the scope of the study.

Potential Threats to Validity

There were a few potential threats to validity. First, the use of certified performance technologists as a sampling frame introduced the potential for sample selection bias. Re-conceptualizing the target population in this manner was supported by the use of similar practices in studies performed by Guerra (2001) and Van Tiem (2004). Still, the reader is cautioned against drawing inferences about performance improvement practitioners, as a whole. Generalizations to performance improvement *professionals* may be less circumspect.

As always, subjects' self-reports should be regarded with some skepticism and not necessarily as matters of fact (Des Jarlais, 1998; Jansen, van de Looij-Jansen, Ferreira, de Wilde, & Brug, 2006; Molenaar, Van Ameijden, Grobbee, & Numans, 2007). As will be seen in Chapter 3, subjects responded to scenarios rather than being observed in a natural, real-life intervention selection decision. This may introduce some potential threats about the generalizability to actual practice. However, the scenarios were developed along guidelines for creating situational judgment tests and low fidelity simulations, both of which have been shown to be predictive of on-the-job performance (McDaniel & Nguyen, 2001; Motowidlo, Dunnette, & Carter, 1990; Motowidlo & Tippins, 1993; Weekley & Jones, 1999). Furthermore, other advantages (e.g., controlled study of key variables and observation of a larger sample size than would be garnered by direct observation), outweighed the potential threats introduced by using a scenario-based approach.

While some might see the quasi-experimental design as a potential threat, it should be noted that some issue has been raised about the use of control groups and comparison of group means for performance improvement studies and that repeated measure designs may, in fact, yield more useful findings. Along these lines, Brethower (2000) noted that:

> Statistical power comes from a few individuals rather than a few observations from individuals. With multiple observations, changes can be statistically significant even though there is only one subject, a fact about statistics that is sometimes lost on researchers in general who equate the statistics n with the number of subjects rather than number of observations." (p. 41)

Of course, within-subjects designs may introduce contextual effects such as practice or carryover (Greenwald, 1976). The seventh research question was developed with the intent of addressing concerns about practice effects. These results will be presented and discussed in Chapters 4 and 5.

CHAPTER 2 LITERATURE REVIEW

As a means for elucidating the study's key variables, the chapter begins with an introduction to the theoretical basis for the study in Normative Decision Theory, including subjective probability, Bayes' theorem, and Bayesian confirmation theory. It then turns to Behavioral Decision Theory and the cognitive biases that illustrate deviations from Bayes' rule for decision-making. Discussion then turns to the field of performance improvement and its historical development to a broader recognition of multiple performance interventions. This section closes with a review of various guidelines for intervention selection.

At that point, Chapter 2 will take a different tack and explore the study's major variables. These variables are investigated both from a conceptual standpoint and on the basis of existing empirical research. However, as will be illustrated in the subsequent pages, research on evidence and PI professionals' intervention selection decisions is both scarce and problematic.

Probability Theory

At almost every turn in our everyday lives, we are faced with problems of probability. For example, "Will I require my umbrella today?" or "Which wine would my dinner party host prefer?" (Jeffrey, 1983). Beyond our mundane day-to-day choices, probability is the very foundation of empirical approaches to social science (Hájek, 2007).

Despite its ubiquity, formal rules of probability were not fully formulated until 1933 when Kolmogorov published *Grundgebriffe der Wahrscheinlichkeitsrechnun,*

which was later translated into English (Kolmogorov, 1956). Skyrms (1966) restated the rules as follows:

1. assign logical truths (i.e., tautologies) such as "a or not a" a probability of 1;
2. assign self-contradictions such as "a and not a" (expressed in logical notation as "a & ~a") a probability of 0;
3. assign logically equivalent statements, such as "a" and "~~a," equal probabilities;
4. if *p* and *q* are mutually exclusive, then the probability of *p* or *q* (logically expressed as "*p* v *q"*) is equal to the probability of *p* plus the probability of *q*. In logical notation, this rule is expressed as Pr (*p* v *q*) = Pr (*p*) + Pr (*q)*;
5. Pr (~p) = 1 – Pr (*p*);
6. whether or not *p* and *q* are mutually exclusive, Pr (*p* v *q*) = Pr (*p*) + Pr (*q)* – Pr (*p & q*);
7. if *p* and *q* are not independent (i.e., conditional), the probability of *p* & *q* is equal to the probability of *p* multiplied by the probability of *q* given *p*, (i.e. "*q* | *p"*). In logical notation, this rule is expressed as Pr (*p & q*) = Pr (*p)* x Pr (*q* | *p*); and
8. if *p* and *q* are independent, the Pr (*p & q*) = Pr (*p*) x Pr (*q)* (pp. 111-123).

This probability calculus has become the measure for an admissible theory of probability (Hájek, 2007), of which there are a number of contenders (Howson, 1995; Parmigiani & Inoue, 2009). These interpretations can be classified into five categories: (a) classical, (b) logical, (c) frequency, (d) propensity, and (e) subjective interpretations

(Hájek, 2007). As an introduction, Table 2 summarizes the major features of each account of probability.

The subjective interpretation is inherent in this study. It takes probability as "degree of belief" (Ramsey, 1926), the main premise being that far from being of the all-or-nothing sort, beliefs come in degrees. For rational decision-makers, these degrees of belief—or *partial beliefs* or *credences* as they are also called—obey the probability calculus (Erikkson & Hájek, 2007). Put plainly, the subjective degrees of belief held by a decision-maker are themselves, probabilities.

Table 2

Interpretations of Probability

Interpretation	Definition	Feature(s)
Classical	$\frac{\text{favorable cases}}{\text{equipossible cases}}$	• Earliest account • The notion of chance or "equipossibility"
Logical	$\frac{\text{favorable cases}}{\text{possible cases}}$	• Unequal distribution of possibility • Focused on the confirming properties conferred from evidence to a hypothesis
Frequency	$\frac{\text{favorable cases}}{\text{finite actual cases}}$	• Equipossibility • Samples
Propensity	$\frac{\text{favorable cases}}{\text{actual cases over the 'long run'}}$	• Long run tendencies
Subjective	degree of belief or confidence	• Many interpretations • Related to notions of preference and betting

Note. Based on *Interpretations of Probability* by Hájek (2007)

Subjective Probability

Predominant thinking on subjective probability comes from Frank Ramsey's seminal work, *Truth and Probability* (1926). Perhaps due to the primacy of his thinking

on subjective probability, Ramsey recognized the ambiguity and difficulty involved in measuring degree of belief:

> The subject of our inquiry is the logic of partial belief, and I do not think we can carry it far unless we have at least an approximate notion of what partial belief is, and how, if at all, it can be measured. It will not be very enlightening to be told that in such circumstances it would be rational to believe a proposition to the extent of 2/3, unless we know what sort of belief in it that means." (p. 166)

Ramsey proposed what he called a "purely psychological" method for measuring belief that could assign magnitudes to beliefs, essentially placing them on an ordinal scale. While the ordinal ranking of beliefs is critical, it does not represent the entire requirement of measuring partial beliefs. There must also be some meaningful ascription of numerical values to beliefs:

> We can of course easily explain that we denote full belief by 1, full belief in the contradictory by zero, and equal beliefs in the proposition and its contradictory by 1/2. But it is not so easy to say what is meant by a belief 1/2 of certainty, or a belief in the proposition being twice as strong as that in its contradictory. This is the harder part of the task, but it is absolutely necessary; for we do calculate numerical probabilities, and if they are to correspond to degrees of belief we must discover some definite way of attaching numbers to degrees of belief. (p. 168)

There are generally two methods for measuring subjective probability. The first involves direct elicitation of an assessment of likelihood, while the second approach indirectly measures subjective probability through preference for a course of action or bet

(Hampton, Moore, & Thomas, 1973; Hogarth, 1975; Wallsten & Budescu, 1983). The latter betting interpretation is grounded in the relationship between probability and preference. Fishburn (1986) describes this relationship as follows:

> …to say that you regard rain as more probable tomorrow than shine, or that you believe the pound sterling is more likely to fall than rise against the dollar next year means roughly that you would rather bet on the first-named event for a valuable prize that you will receive if your chosen event obtains." (p. 335)

Ramsey found the long-standing practice of measuring probability via betting behavior sound. In fact, according to Erikkson and Hájek (2007), such a betting interpretation has played a key role in other subjectivist views of probability as well. Bruno de Finetti held a view of degree of belief that essentially equated degrees of belief with betting prices, while Richard Jeffrey held them as correlated with each other (p. 191).

From an empirical perspective, there is "reasonable consistency" between direct likelihood elicitation and indirect betting behavior methods, but inconsistencies appear to be a function of the assessors' relative experience with probabilistic reasoning. These inconsistencies are somewhat reduced by practice (Hogarth, 1975, p. 279). Wallsten and Budescu's review relating to the inherent issues of reliability, validity, and scaling involved in measuring subjective probability found that in those who were naïve in regard to probability assessment, there were differences in results for direct likelihood and indirect betting methods of measurement. Moreover betting approaches resulted in more inconsistent estimates (Wallsten & Budescu, 1983). Furthermore, given mixed support for individual differences in risk preferences and aversion, measuring probability

through one's propensity to accept a wager remains problematic, (cf. Gertner 1993; Metrick, 1995; and Hersch, 1997) on risk aversion in game environments. Measurement of probability in the present study will be discussed further in the section on variable specifications in Chapter 3.

At this point, it is important to note the role of probability in normative guidelines for decision-making, especially its relationship to expected utility theory, which asserts that in order to decide between two options, one need only determine the perceived likelihood of each outcome (*p)*, the desirability of or preference for that outcome (*d*), multiply them together to obtain the expected utility for each act, and then compare expected utilities, the act with the highest utility being the optimal choice (Jeffery, 1983). Ultimately, the subjective interpretation is a functionalist account of degrees of belief, specifically, "they are whatever fills the role of being multiplied with utilities in the expected utility representation" (Erikkson & Hájek, 2007). This perspective hints at a practical strength of the subjective interpretation of probability, if degrees of belief are "whatever" it is that interacts with utility in the expected utility framework, then they make for neat calculations in decision problems and are easily applied as a guide for decision-making (Hájek, 2007).

Having examined the broader framework for subjective probability, this chapter now turns to the particulars of Bayes' theorem and its connection to confirmation theory.

Bayes' Theorem

A version of Bayes' theorem was first presented in a posthumous paper by Reverend Thomas Bayes (1763). The paper, titled *An Essay Towards Solving a Problem*

in the Doctrine of Chances, was bequeathed to a former colleague, Richard Price (Earman, 1992; Stigler, 1982). Despite Price's recognition of the potential significance of the work for inductive reasoning, Bayes' paper and the theorem were largely ignored until the twentieth century work of Karl Pearson, Ronald Fisher, Harold Jeffreys, and Frank Ramsey (Bolstad, 2007; Earman, 1992; Psychology, 2006). Interestingly, Stigler (1975) posits that Pierre-Simon Laplace's similar work on inductive reasoning during the late eighteenth century was performed in absence of prior knowledge of Bayes' work. Whether or not this is the case, Bayes' preeminence is clearly established in the discipline as the theorem, rule, movement (i.e., Bayesianism), and its followers (i.e., Bayesians) all bear his name. So do the many areas and approaches that draw on his work, as in Bayesian confirmation theory, Bayesian decision theory, Bayesian statistics, Bayesian analysis, Bayesian networks, Bayesian computation, etc.

As presented in Chapter 1, the most basic statement of Bayes' theorem is:

$$\Pr(h|e) = \frac{\Pr(e|h) \times \Pr(h)}{\Pr(e)}$$

Bayes' theorem is a noncontroversial consequence of the probability calculus in general and the axioms related to conditional probability and logical equivalence in particular (Earman, 1992; Salmon, 1967; Skyrms, 1986). It is often used in connection with claims about the nature of evidential support and theory confirmation.

First, is the evidence proportionism principle, namely: an individual's degree of belief in a particular hypothesis ought to be a function of the body of evidence relevant to it (Joyce, 2003). From a temporal perspective, probabilities may be distinguished from

one another on the basis of whether they precede or follow the acquisition of new evidence, as illustrated in Figure 3.

Salmon (1967) stated that since prior probabilities temporally precede data collection they are concerned with the plausibility of a hypothesis. Evans et al. (2002) discussed prior probability as a base rate that "refers to the degree of belief that we have about the hypotheses before considering any specific new piece of evidence" (p. 179).

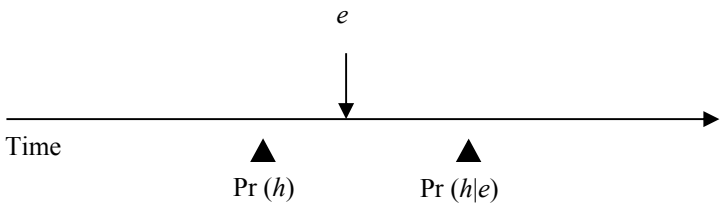


Figure 3. Temporal relationship of probabilities to the introduction of evidence.

Note: Pr (*h*) represents prior probability, *e* represents introduction of evidence and Pr (*h*|*e*) represents posterior probability.

Similarly, Blackburn (2005) discusses prior probability as "[t]he probability assigned to a hypothesis or event before a piece of evidence emerges" (p. 292). By extension, posterior probability occurs after additional evidence is acquired.

While Bayes' theorem is a simple consequence of the probability axioms, Bayes' rule provides a normative directive for calculating an optimal posterior probability, once new evidence obtained: Pr (*h*|*e*). This point has a key implication for the measurement of the strength of evidence. Theories are confirmed—or disconfirmed—incrementally. Where evidence, *e,* may be said to confirm a hypothesis, *h*, if and only if Pr (*h*|*e*) is greater than Pr (*h*). By extension, the opposite is true, evidence may be said to disconfirm a hypothesis, if and only if Pr (*h*|*e*) is less than Pr (*h*) (Earman, 1992). Therefore, by

comparing differences between posterior and prior probabilities we can measure the strength of the evidence:

> When scientists (or ordinary folk) say that E supports or confirms H what they generally mean is that learning of E's truth will increase the total amount of evidence for H's truth. Since subjectivists characterize total evidence in terms of subjective probabilities or odds, they analyze incremental evidence in terms of changes in these quantities. On such views, the simplest way to characterize the strength of incremental evidence is by making ordinal comparisons of conditional and unconditional probabilities or odds." (Joyce, 2003, section 3)

In laymen's terms, all confirmatory evidence is not equal and it comes in varying degrees, depending on the extent to which it increases the likelihood that a hypothesis is true.

Theoretical Alternatives and Practical Problems

There are many parallels between subjective probability, the Bayesian account, and intuitions about decision-making and theory confirmation. Still, there are a number of obstacles, especially in contrast to other views of probability, namely the frequentist paradigm.

By and large, one of the most commonly recognized approaches to inference is the frequentist method (Bland, 1998; Bayarri & Berger, 2004). As already discussed, the subjectivist view takes an iterative approach to theory confirmation and decision-making, where probability is conditionalized based on new evidence. In contrast, frequentist probability is defined as a proportion of the number of "favorable" cases to the number of

all observed cases (i.e., the summation of actual favorable and non-favorable cases). In the late nineteenth century, Venn was one of the earliest to propose a frequentist probability theory, on the basis of finite proportions and various other views (e.g. infinite, relative) have been put forth in response to theoretical problems for frequentist probabilities (Hájek, 2007). While Bayesian methods prevailed in the prior century, frequentist methods greatly overshadowed them in the twentieth century (Efron, 1986, 2005; Wagenmakers, Lee, Lodewyckx, & Iverson, 2008).

In many respects, the ascendance of the frequentist view is closely linked with the fact that rapid development within the field of statistics outstripped the computational capabilities of the day. The suitability of each paradigm's use by novices or scholars is a contentious topic and pedagogy is also likely at play in the lasting influence of frequentism. For example, Bayarri and Berger (2004) note that they "…would probably argue that Bayesian statistics… should be the type of statistics that is taught to the masses, with frequentist statistics being taught primarily to advanced statisticians…" (p. 59).

Empirical Problems and Cognitive Biases

Up to this point, discussion has mainly focused on theoretical challenges to subjective probability and Bayes' rule for updating probability based on evidence. However, its utility as a behavioral theory is circumspect:

> In the statistical literature probability updating is well described by Bayes' theorem. Behavioural scientists (notably experimental psychologists and economists) however, found that Bayes' rule is not necessarily a well-functioning

> descriptive and predictive device….Apparently, the rule is often not (or wrongly) applied in the process of updating. The result is that reported posteriors deviate from the normative outcome represented by Bayes' rule, which we interpret as errors decision makers make in the updating process (Ouwersloot, Nijkamp, & Rietveld, 1998).

These behavioral deviations are attributed to the use of heuristics. Heuristics are rules of thumb that are employed to circumnavigate the complexity involved in updating beliefs on the basis of Bayes' theorem. When used, these heuristics result in systematic errors or effects, known as cognitive biases (Tversky & Kahneman, 1974). The following section summarizes several cognitive biases that are relevant to the design of the present study.

Base rate neglect. According to Bayes' theorem, prior probabilities and the evidence ought to carry equal weight, yet errors are often made in combining them. The phenomenon is known as base rate neglect, where greater credence is given to specific evidence-at-hand than to prior probabilities (Bar-Hillel, 1980).

Kahneman and Tversky (1973) first demonstrated base rate neglect through an experiment where subjects received the frequencies of engineers (n=30) and lawyers (n=70) within a group of 100 professionals. Then, subjects reviewed short, randomly-selected descriptions of individuals and were asked to estimate the probability that that individual was a lawyer or an engineer. A second group performed the same task—with the same individual descriptions—only the proportion of lawyers and engineers were reversed. Despite this difference, the probabilities in both groups were similar. Following these assessments, subjects provided a probability of being a lawyer or engineer on the

assumption of having no information. These probabilities tended to follow the proportions provided in the base rates. Most surprisingly, subjects reacted to descriptions that were non-informative regarding profession differently, estimating probabilities at or around 50%.

Since the problem was first introduced, research into base rate neglect has focused less on whether it exists and more upon the contextual circumstances in which it occurs (Gigerenzer, Hell, & Blank, 1988). Early on, the extent that individuals were stereotypical or representative of a category was demonstrated as a key factor (Kahneman & Tversky, 1973). Over time, learned associations between stimulus and outcome (Goodie & Fantino, 1996), and the simple availability of case-specific information (Tversky & Kahneman, 1982) were shown to contribute to base rate neglect. On the other hand, base rate neglect can be mitigated by making base rates appear equally relevant with case-specific information (Bar-Hillel, 1980). Others have found that the effect may be erased when base rates are highly diagnostic or framed in a frequentist approach (Cosmides & Tooby, 1996; Koehler, 1996). Evans et al. (2002) found that base rates could even be overvalued by asking subjects to provide the probability assessments themselves.

Confirmation bias. This reasoning error involves "…the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand" (Nickerson, 1998). As suggested by this definition, there are two major streams of research in this area: (a) information search and (b) interpretation of evidence.

A number of factors have been shown to influence confirmation bias in information search, including the extent to which the decision-making process focuses on discovery rather than a goal (Jonas, 2008), expectation for discussing one's viewpoint (Mojzisch, 2008), and use of graphical representations of the body of evidence (Cook, 2008). In fact, confirmation bias is so pervasive as to occasionally result in false memories or the perception of supportive evidence where it does not exist (Tschan et al., 2009).

With regard to the interpretation and weighting of evidence, Edwards (1968) as cited in C.R. Chapman (1973) noted that "[Decision-makers] treat data that *support* their preference (confirming data) differently than equally diagnostic data that contradict their preference (infirming data)" (p. 270). For example, Koehler (1993) studied 297 advanced graduate students in various areas of science. Each subject received a 20- to 35-page booklet related to two fictional scientific hypotheses. Those in the experimental groups read a two-page summary of each hypothesis and then a relevant research report. Those in the control group were not presented with a background summary. The purpose of the summary document was to induce a prior belief about the correctness of each hypothesis, indicated on a scale from zero (*very unlikely*) to 100 (*very likely*). Subjects then reviewed two research reports; these reports were either of a high or low quality. Those in the experimental groups also received pages summarizing the results and discussion sections for the research reports. Following that, subjects were asked to respond to a series of analytical and evaluative questions concerned with both content-specific and general judgments about the quality, clarity, and relevance of the reports. Finally, subjects

reported their demographics and answered a series of questions about the extent to which their beliefs were and should have been influenced by the research reports. Overall, subjects rated reports of a higher quality when they agreed with their prior beliefs. There was an agreement main effect and an interaction between agreement and strength of prior belief. For the content specific judgments, marginal agreement effects were found (pp. 37-38). Most (64%) believed that their judgments about the methodological quality of the research reports were not influenced by the extent to which the findings agreed with their prior beliefs and a great majority (83%) supported that methodological quality assessments *should not* be dependent upon their level of agreement. There were no significant differences between the experimental groups related to these beliefs, however "[t]hose who believed that the outcome of the study did not influence their quality judgments were actually influenced by the outcome as much as those who admitted some probable influence" (p. 39). So, introspection on whether one is influenced by the amenability of acquired information is not enough to counter the confirmation effect.

In fact, confirmation bias is highly situational. For example, Wright (1974) studied 210 male undergraduate business students who were assigned to groups that varied in the amount of time pressure and distraction. Significant differences in weighting evidence communicating negative outcomes were noted and moderate effects were also documented for the group who was moderately distracted (p. 559). Yaniv and Milyavsky (2007) focused on agreement and accuracy of advice provided and their influence on final probability estimates for a task related to historical facts. The experiment followed these general procedures: an initial phase where subjects provided estimates in response

to a set of historical questions and a second phase where they were presented with the same questions, constructed estimates from advisors, and then asked to provide a new estimate. This was accomplished via a within-subject factor with four levels configuration: when both pieces of advice point away from truth, a 20% loss in accuracy was observed; when both pieces of advice point towards truth, a 15% gain in accuracy was observed; when the distant piece of advice points towards truth and the near advice points away from truth, a 3% gain in accuracy was observed; when the near piece of advice points towards truth and distant advice points away from truth, a 1% loss in accuracy was observed. Furthermore, one-way ANOVA indicated significant differences among the conditions. Specifically the first two conditions differed from each other. Conditions three and four did not differ from each other, but did differ from both conditions one and two. This finding leads the authors to conclude: "Clearly, good advice helps decision-makers, while poor advice leads them astray. Gains are a function of the quality of advice as much as of the revision rules that one uses." (p. 115). Another experiment reported in this study targeted the integration of initial opinions and advice and changes in judgment accuracy. This was accomplished by examining both the accuracy of subjects' judgments and the fit of various belief revision rules with changes in these judgments. In this experiment, the authors found that estimates improved approximately 27% after receiving two pieces of advice, 28% after getting four, and 33% after eight. Thus, from an accuracy standpoint, the benefits of acquiring additional advice diminish fairly quickly. Furthermore, ignoring distant opinions may not always be detrimental, as indicated by the higher levels of accuracy for egocentric approaches to

pruning advice. Despite these intriguing positions, the authors note that the findings are limited by the limited scope of the study to quantitative advice and recommend further study of other types of advice, including probability, preference, and arguments (p. 119).

Rassin (2008) found only marginal support for individual differences in confirmation bias. Using the ten-item Confirmation Inventory, Rassin assessed the confirmation tendencies of 95 undergraduate students. Next, these same subjects were presented descriptions for five scenarios, in turn, and asked to indicate a preferred line of action in response to each situation. A composite variable was constructed that corresponded to the number of scenarios to which the subject's response was confirmatory of the information presented; this composite variable demonstrated a very a low level of reliability (Cronbach's α = .03) that was significantly correlated with performance on the Confirmation Inventory (r = .44). Furthermore, the Confirmation Inventory scores for those who chose confirmatory actions were compared with those who chose non-confirmatory actions on each scenario and p-values were reported at .16, .03, .001, .006, and .93 respectively. Based upon these findings, Rassin concluded that there are indeed individual differences in confirmation bias (as measured by the Confirmation Inventory), confirmation approaches are favored, but that confirmation bias is highly contextual: "[S]ituation dependence is quite strong, in that individuals do not make confirmatory (or nonconfirmatory) choices reliably" (p. 92).

Escalation of commitment. From a normative perspective, decision-makers ought to proportion their beliefs to the evidence they have. By extension, it is reasonable to expect changes of mind regarding a preferred course of action if enough compelling

counter-evidence is obtained. However, a variety of studies demonstrate the opposite, namely that commitment to a course of action may be escalated despite negative consequences (Staw, 1976)**.** The phenomenon was first studied in terms of failed political decisions and investment decisions, where it is sometimes termed the *sunk cost fallacy.* The former tended to focus on groupthink processes (cf. Janis, 1972; Kramer, 1998; and Raven, 1998), while the latter emphasized a multi-factored model of commitment escalation, including motivation to justify previous choices, consistency norms, perceived likelihood of future outcomes, and the perceived value of future outcomes (Staw, 1981). More recently, Sleesman et al.'s (2012) meta-analysis of 166 studies of human escalations of failed actions showed significant effects for 14 of the 16 independent variables examined, as shown in Table 3:

Table 3

Main Determinants of Escalated Commitment

Decreased escalation	Increased escalation
• Risky decision • Presence of opportunity costs • Information acquisition • Anticipated regret • Positive information framing	• Uncertainty of decision • Positive performance trends • Expressed preference for initial choice • Sunk costs • Time investment • Decision-maker experience/expertise • Self-efficacy or confidence • Personal responsibility for decision • Ego threat • Proximity to project completion • Group cohesiveness • Agency problems

Others, (e.g., Tsai and Young, 2010) examined the relationship of negative emotions, such as anger or fear, and escalation of commitment. Higher levels of

escalation were observed with anger than fear, but these effects were primarily a function of lower risk perceptions.

Summary

This section reviewed the core elements of subjective probability and Bayes' theorem. Although empirical research on cognitive biases indicates Bayes' theorem is circumspect as a descriptive decision theory, it remains as a viable framework for studying decisions in general and of performance improvement in particular, mainly in its connection to confirmation theory and evidential support. On this view, the following section discusses the study's basis in performance improvement.

Performance Improvement

The field of performance improvement is also commonly known as "human performance technology" (Pershing, 2006), "human performance improvement" (Stolovitch, 2007; Stolovitch & Keeps, 2006), "performance consulting" (Robinson & Robinson, 1996; Rummler, 2004), "performance engineering" (Gilbert, 1978) and various alternatives that emphasize the humanistic or technological elements to varying degrees (c.f. Hybert, 2003; and Mager, 1992). Yet the theme underlying all of these labels is the notion of "performance" (Guerra, 2001b).

The impetus for a field of performance improvement was mainly derived from problems of an educational nature (Brethower, 2008; Rummler, 2007) and it has often been linked with instructional systems design (Guerra, 2001b; O'Driscoll, 2003; Reiser, 2001; Rosenberg, Coscarelli, & Smith Hutchinson, 1999; Tosti & Kaufman, 2007) and the programmed instruction movement (Buxton, 1982; O'Driscoll, 2003; Shoemaker,

1972). However, as educational results were realized but not translated into practice in the workplace, the 1970s ushered in a shifted focus from training and instruction to performance (Buxton, 1982; Rummler, 1982; Tosti, 2005). Practitioners began to frame their work in real-world problems, which led to the recognition that problems were caused by a variety of factors and, therefore, could not be solved solely by training (Buxton, 1982; Ruckdeschel, Yarter, Riveccio, Cortes, & Cookson, 1998; Rummler, 1982).

As practitioners in the field began to recognize that performance problems could be solved by many different types of means, they began to refer to these means as "interventions." Westgaard (1996b) stated that he began to use the term in publications around 1980 and review of article titles in NSPI's journals reveals several references to instructional "interventions" around that time (e.g., Buxton, 1984; Davis, Latham, & Pitts, 1985; Lindsey & Cheek, 1986; Schwen, Leitzman, Misanchuk, & Foshay, 1979). Westgaard (1996) implied that the term may have been borrowed from medicine and psychology, where it existed previously, but Schwen et al.'s (1979) usage indicated an orientation to intervening in social problems. This latter explanation fits well with the notion of an intervention as the means by which performance problems could be solved by resolving discrepancies between current and desired results (Buxton, 1984; Gilbert, 1978; "Whatever happened to what's its name, programmed instruction? Or a front-end/rear-end analysis of 'fuzzies' as organizational goals," 1973). This dual-orientation of intervention to both the present and what should be in the future endures in current definitions of the term. Nickols (2005) referred to interventions as "purposeful action" (p.

9), while Farrington and Clark (2000) defined them as "the means we use to eliminate barriers to achieving worthy goals." (p. 6).

Commonly, these interventions are organized into classes. Several of these classification schemes are summarized in Appendix A. As the field has developed, intervention classification systems have played a variety of roles in the field. These roles range from documenting historical development, (Hill & Brethower, 1997; Van Tiem et al., 2001), to establishing the boundaries of the field (Hutchison, Stein, & Carleton, 1996), and also as a means for measuring professional expertise (Hutchison et al., 1996; C. S. Hutchison & Stein, 1998; Van Tiem, 2004). Still others describe intervention classification systems as mechanisms for managing responsibility and interventions selection. Harless, as cited in Langdon (1997b), suggested that [Analyzing the subclasses of interventions]"…gives the technologist a handle on the interventions. Then with this taxonomy, the technologist does not necessarily need to know how to develop each, but rather find those who can develop the appropriate intervention within or outside the organization." (p. 37). Although Harless discussed intervention classification schemata as taxonomies, it is important to note that they would be more accurately described as "typologies." Both terms deal with classification, but unlike the latter, typologies are based upon a priori distinctions rather than empirical observations (Sanchez, 1993, p. 75). Therefore, typologies depend largely on a conceptual analysis of the characteristics of the objects being classified rather than findings from research. This point is especially problematic as intervention typologies often play a role in practices that are

recommended for choosing interventions as will be seen in the following review of intervention selection.

Intervention Selection Models

The initial phases of problem identification that lead up to the intervention selection decision are characterized in a variety of ways (Sleezer, 1992; Watkins, Leigh, Platt, & Kaufman, 1998). One conception of this process is "performance analysis" (Sleezer, 1992). Performance analysis provides a framework for aligning the remainder of the activities involved in performance improvement; it involves the identification of both desired and actual levels for results and the quantification of this discrepancy, or "performance gap," between them (Grant & Moseley, 1999; Jonassen, 1989; Kastigar, 1991; Van Tiem et al., 2000). Rummler and Brache (1995), Brethower (1982), and others have stressed that these activities require explication of the organizational system itself.

Sometimes activities related to the determination of causes, generation, and subsequent selection of possible solutions are included within performance analysis (cf. Guerra-López, 2007; Jonassen, 1989). However, many models separate this activity into a separate phase known as "cause analysis" or "intervention selection" (Van Tiem et al., 2000). Additionally, it is often emphasized that performance discrepancies are gaps in results and that these gaps in results represent "needs" (Kaufman, 1985). In fact, many authors advocate an additional step within performance analysis to prioritize needs, known as "needs analysis" (Guerra, 2001b; Kaufman & Valentine, 1989; Watkins & Kaufman, 1996).

Van Tiem et al. (2000) stated that intervention selection begins after the organization, its environment, discrepancies in performance, and the conditions that have potentially caused it have been analyzed. In fact, the authors closely linked intervention selection with cause analysis, noting "practitioner[s] must select the interventions that work best according to the problems identified" (p. 63). This viewpoint, that the appropriateness of an intervention is primarily dependent upon its likelihood for resolving the causes of a performance discrepancy, is shared by many others (e.g., Brown, 1986; Darabi, 2003; and Herem, 1979).

The following section discusses guides for intervention selection.

Cause-based. The most prevailing view of intervention selection is that it follows from having determined the cause for the performance discrepancy. For example, 63% of respondents to Rossett and Tobias' (1999) survey reported that their organizations typically selected interventions based on the data and findings of a cause analysis. An early example of a cause-based approach, Bullock's (1973) described a procedural model that clearly linked the selection of an intervention to the types of causes of the performance discrepancy. This model included phases of (a) problem identification, (b) problem definition, (c) determining solution objectives, (d) comparing the objectives with the actual performance, (e) hypothesizing causes of the problem (f) exploring possible solutions based on likely cause(s), (g) implementing solutions, and (h) evaluating the resulting changes.

Shortly later, Gilbert (1978) distinguished between the ultimate cause for performance discrepancies (i.e., deficiencies in the management system) and immediate

causes for a performance problem which may occur in a performer's behavior or in the environment (or both), namely (a) data, (b) instruments, (c) incentives, (d) knowledge, (e) capacity, and (f) motives. Although causes for performance deficiencies may reside in each of these areas, diagnosis of deficiencies—and therefore selecting the interventions that are associated with them—ought to occur in this sequential order: (1) data, (2) instruments, (3) incentives, (4) knowledge, (5) capacity, and (6) motives. This is not due to decreased importance among the areas but as one progresses, the possibility of leveraging performance change decreases.

A somewhat similar mindset was established by Mager and Pipe (1997). Their flowchart guides a sequential process of intervention selection, where if certain criteria are met by an observation of the performance deficiency, analysis stops and a related intervention is selected. Generally speaking, this sequence addresses the following questions:

1. What is the performance problem?
2. Is this problem worth solving?
3. Can a "fast fix" such as clarifying expectations, providing addition resources, or communicating feedback be applied?
4. Are the consequences aligned with desired performance?
5. Do performers already know how to perform the related tasks?
6. Are there additional indications of what can be done (e.g. simplifying tasks, other obstacles, performer's potential for change)?

7. The best solution is selected based on the answers to these questions and cost calculations.

Darabi (2003) also put forth an explicitly cause-focused model for intervention selection including phases of cause identification, prioritization of causes by impact on the organization at multiple levels, classification of causes, and selection and recommendation of interventions based on this classification. Paralleling cause-related approaches to intervention selection, Watkins (2007b), Watkins and Wedman (2003), and Wedman and Graham (1998) included the following classes of interventions: (a) skills and knowledge; (b) motivation and self-concept; (c) performance capacity; (d) expectations and feedback; (e) tools, environment and processes; (f) rewards, recognitions, and incentives; and (g) strategic, tactical, and operational directions. However, rather than the taking a troubleshooting approach, this model emphasizes alignment, consideration of how interventions might interact with one another, and the elements that are required in a performance system in order to be successful (Watkins, 2007b). Despite their prevalence, cause-based guidelines for intervention selection have their limitations, especially insofar as their reliance on a logical relationship between an intervention and causes for a performance deficiency, as well as an intervention's causal relationship with the desired level of performance.

Change-focused. Langdon (2003) cited that "[c]ause analysis is imprecise and does not necessarily take into account the change in performance; rather, it encourages jumping to a solution while masquerading as a scientific step" (p. 8). Instead, he advocated performing a change-of-state analysis. Change-of-state analysis presents a

unique approach to intervention selection that considers the type of change that is required (Langdon, 2003; Langdon et al., 1999).

Rather than deciding on possible interventions in relation to a particular type of cause, the practitioner first determines the change-of-state that is required. These changes of state may involve (a) extinguishing, (b) maintaining, (c) establishing, or (d) improving performance. This classification is an intermediary step that precedes intervention selection, which is derived from a matrix of particular types of interventions that are shown to be effective for particular changes of state (Langdon, 2003; Langdon et al, 1999). There are clearly parallels between the states targeted by Langdon and the major facets of classical conditioning, (e.g., extinction and reinforcement, R. C. Richey, 1986). Therefore, a notable issue for this approach is Langdon's use of "performance," when he really seems to be targeting *behavior.* Equally problematic is the usage of "improvement" as a change state. Here, Langdon appears to equate improvement with an increase, when it is quite possible to decrease an occurrence (e.g., error) which may result in an overall improvement in performance. In addition to this oversight, change-of-state analysis ultimately falls prey to one of the key problems for cause-focused guidelines for intervention selection: it relies on a logical opinion about the effectiveness of particular interventions for changing the performance situation at hand. As this typology was developed by a group of self-touted experts, it may be well-informed, but without further support, the suitability of interventions included in it remains simply a matter of opinion.

Outcome-oriented. An alternative method for systematically selecting interventions involves the development of performance requirements (R. Kaufman,

Oakley-Browne, Watkins, & Leigh, 2003; Svenson, 2006). A requirement bridges the processes of analysis and design; it is a "technical statement about some attribute of the solution that can be validated and tested during design, development, and implementation" (Svenson, 2006, pp. 223-224). Kaufman et al. (2003) also advocated establishing a set of requirements before selecting an intervention. Following this process, they recommended performing a methods-means analysis to identify and compare various intervention alternatives. Performing a methods-means analysis involves asking and answering several questions:

- Should strategies be used to generate possible solutions?
- How does or should an Ideal Vision influence the selection of an intervention?
- How can comparisons between alternatives be made?
- What are the advantages and disadvantages of each?
- Should the feasibility of solutions be analyzed systematically?
- Are ready-made solutions available?
- What constraints exist for the solution that can be selected?
- Could brainstorming help identify solutions?
- Does the team have the skills to recognize the best solution?
- Do we require assistance with the selection decision?
- In light of the current paradigm, is the desired objective possible?
- Will different ways of doing things have to be learned?
- What solutions have worked (or failed) before?
- Are there ethical issues to consider in the selection decision?

- Who can approve the decision?
- What are the potential risks for the possible solutions?
- What are the likely consequences for each decision?
- Are any of the potential risks unacceptable? (pp. 177 -178)

An extension of this involves examining multiple relevant important consequences and identification of possible intervention means for obtaining them (Kaufman et al., 1997; Muir, Watkins, Kaufman, & Leigh, 1998). A notable example of this approach, Stolovitch and Keeps' Performance Intervention Selection Rating tool (2008) considered a set of possible solutions or actions in terms of appropriateness, costs, feasibility, and client acceptability. Each intervention should be rated according to each criterion on a four-point scale. These ratings would then be summed, which provided a basis for an ordinal ranking, and the decision would be made to retain or eliminate the intervention from future consideration. The process that the Performance Intervention Selection Rating tool facilitates is similar in approach to multi-attribute utility analysis (Watkins, personal communication). Given the complexity of multi-attribute utility analysis, Stolovitch and Keeps surprisingly asserted that when compared to the front-end analysis processes, selecting a solution is the "easy part" of performance improvement (p. 149). Despite this overstatement, the Performance Intervention Selection Rating tool remains a promising intervention selection tool. Along with change-of-state analysis, method-means analysis, and performance requirements, it represents a paradigmatic shift away from the basis of intervention selection in the cause for a performance discrepancy.

Despite the plethora of guidelines that exist for systematic intervention selection, actual practice is less methodical than recommended (Langdon, 1992, 1997a; Langdon & Whiteside, 1997). Lovelady's (1984) study of nine consultants noted "In their accounts of their work and the methods used, consultants tended to offer vague, unformed reasons as to why they had chosen certain methods. In general, they freely described what had occurred in a project, but rarely explained why they had chosen particular methods" (p. 7). In fact, intervention selection may not frequently be a decision that is made by PI professionals: Guerra (2001b) and Guerra (2003) measured professionals' beliefs about discrepancies between how often various PI activities ought to be performed and how frequently they actually are. Here, the largest discrepancies occurred for tasks relating to determining what types of intervention are required and reviewing analysis results before interventions are designed. Similar results are seen with instructional designers, who rarely develop and weigh alternative solutions at a broad level; rather they tend to examine various options within a specific solution (Visscher-Voerman & Gustafson, 2004). In fact, the most commonly reported reason for not performing design activities was that decisions had already been made (Mann, 1996; Tessmer & Wedman, 1992; Winer, Vasquez-Abad, & Tessmer, 1994).

At this point, Chapter 2 turns from literature related to intervention selection and begins to explore the conceptual and theoretical basis for the major variables involved in this study. Each variable is discussed in turn and the major issues are addressed in the summary section.

Nature of Evidence

Practitioners should use multiple methods to gather information to aid in the intervention selection decision. Beer (1996) discouraged "mono-methods," noting that they present the risk for putting all of the intervention "eggs in one basket" (p. 79). As an alternative, authors have advocated the importance of knowing and using a variety of data collection techniques (cf. Marrelli, 2010). Within these arguments, triangulation is usually advocated. Triangulation uses "multiple observations (measurements) of different objects" to approximate truth (Baker, Grubbs, & Ahern, 1990, p. 27). While it is often a part of discussions of using multiple data collection methods, triangulation emphasizes both the variety and suitability of evidence collected (rather than merely using multiple approaches for collection).

Guerra (2003) recommended the selection of a source of evidence prior to making decisions about what tools ought to be employed in order to access it: "…before you can collect data, you must first determine where it can be found" (p. 27). Thomas (2006) also underscored the importance of considering the source of evidence, rallying practitioners to:

> "…be fully informed about best available evidence, the strengths and limitations of different sources of evidence, and the strengths and limitations of our expertise, as applied to each situational-specific context. Best practices reflect the integration of multiple sources of evidence, everything from research, to practitioner's experiences, to situational context." (p. 10)

Despite this recognition of relevant evidence arising from a variety of sources, a great deal of concern has been expressed that more craft-based sources have become over-represented in PI professionals' decisions, while sources of a scientific origin are under-represented.

Stolovitch (2000) claimed that some beliefs commonly held by PI practitioners are "mythical" when viewed in light of research. Some of these beliefs include intuitive statements related to (a) feedback leading to improved performance, (b) timely feedback being more effective than delayed feedback in improving performance, (c) performance improvement demonstrated during training extending to post-training performance, (d) experts being a good source for procedural knowledge and learning, (e) investments in human capital resulting in lower returns than investments in physical capital, (f) increases in productivity since 1970 due to technological advances, and (g) common sense as an ally to science. Similarly, Farrington and Clark (2000) utilized several case studies to demonstrate that commonly used tactics such as the Myers-Brigg type indicator, delivering training on a computer to increase learning and retention, and interviewing experts to understand expert practice may likely be little more than *snake oil*, meaning that on their face these tactics appear scientifically-based, but in reality they have no effect. Other authors have raised similar concerns about practitioners' over-reliance upon craft-based approaches and sources of evidence including authority (Thomas, 2006), intuition (Langdon, 1997), pre-existing common sense beliefs (Hannum, 2009; Langdon, 1997; Thomas, 2006), innovation over reliability (Sugrue & Stolovitch, 2000) and familiarity (Clark, 2003; Hutchison & Stein, 1998; Langdon, 1997b).

The staunchest criticism of the field's dependence upon these sources comes in a series of works by (Clark & Estes, 1998; Clark & Estes, 2000; Clark & Estes, 2002) Here, the authors distinguished craft-based approaches from scientific methods and identified the sources of evidence from which each draws. Craft-based approaches draw on luck, personal expertise, insight, the experiences of others, and trial-and-error to solve specific problems. On the other hand, scientific inquiry aims at generalized principles of how the world works and is tied to principles, theoretical models, experimentation, and clear specification of problems. Additionally, science affords opportunities for disconfirmation and isolation of key performance factors.

Empirical support for Clark and Estes' assertions that practitioners tend to rely on craft-based approaches is limited: review of the field's primary serial publications and journals dating back to 1962 (i.e., Performance Improvement Quarterly, Performance Improvement, Performance and Instruction, Improving Human Performance Quarterly, Improving Human Performance, NSPI Journal, and the NSPI Newsletter) revealed no studies directly examining the intervention selection process and the role of evidence in it. The lack of the issue's explicit consideration in the empirical literature is further supported by the identification of this issue as research priority in the field (Huglin et al., 2007).

However, a few studies deal with the nature of evidence that is used in professional practice more implicitly. As one indicator of professional practice, reviews of the published literature suggests a lagging emphasis on scientific data. For example, Lindsley as cited in Binder (1995) called into question the reliance of the field's authors

on scientific measurement, finding that fewer than 5% of the tables or figures in the then NSPI publications contained measures of performance results. Binder himself noted that only four out of 60 contributors to the first edition of the *Handbook of Performance Technology* (Stolovitch & Keeps, 1992) shared performance data. More recently, Guerra (2001a) studied the theoretical nature of the field via a content analysis of the practitioner journal, PIJ, between January 1998 and June 1999. Of those articles included for review, only 21% were scientifically-based while 79% were craft-based, meaning that the selection and design of solutions did not result from systematic analysis of a problem or sound research. These findings seem to suggest less reliance in the field on scientific approaches.

Studies of professionals' reported practice provide more compelling data related to the usage of evidence in intervention selection decisions. Rowland (1992) observed four expert and four novice instructional designers as they reviewed materials and developed a content outline for an instructional solution to a hypothetical problem. Although the data analysis protocol revealed heavy emphasis on operational activities, the synthesis and interpretation of the data involved the development of graphical representations that generally illustrated both the types of evidence used and the frequency with which they were consulted. Experts reportedly accessed memories of past experiences as a designer, prior knowledge of templates for how to proceed to understand the problem further, resource materials provided about the problem, and possible results of hypothetical interactions with subject matter experts and other stakeholders. During the generation of solutions, experts recalled templates for solutions and principles of

design from memory. Throughout the entire process, novices relied heavily upon the problem materials. During solution generation, novices also drew upon memories of personal classroom experiences and general reasoning.

In a more recent effort to understand instructional strategy decisions, Christensen and Osguthorpe (2004) surveyed 150 alumni from graduate instructional design programs at five universities. As a part of the study, subjects indicated how frequently they used particular strategies for making decisions about instructional strategies on a 5-point semantic differential scale. These strategies included:

- Generating ideas with others involved in the project
- Comparing the current situation with personal past experiences and then making adaptations in similar cases
- Modifying strategies based on having observed others use them
- Generating ideas individually based on instructional goals
- Considering non-traditional and performance-based options (e.g. job aids, incentives, selection)
- Conferring with subject matter experts for strategy ideas
- Following a template one has developed and used before
- Looking at instruction that has been successful in the past and has similar instructional goals
- Using learning theory or research
- Using prescriptive instructional design theory or research
- Generating ideas with potential learners

- Following a template that has been successfully used by others

Descriptive summaries of these results are included in Table 4:

Table 4

Regularly Used Tactics for Making Instructional Strategy Decisions

Tactic	Percentage of respondents who report regular usage
Generating ideas with others involved in the project	86%
Comparing the current situation with personal past experiences and then making adaptations in similar cases	79%
Modifying strategies based on having observed others use them	74%
Generating ideas individually based on instructional goals	69%
Considering non-traditional and performance-based options (e.g. job aids, incentives, selection)	64%
Conferring with subject matter experts for strategy ideas	58%
Following a template one has developed and used before	58%
Looking at instruction that has been successful in the past and has similar instructional goals	57%
Using learning theory or research	54%
Using prescriptive instructional design theory or research	51%
Generating ideas with potential learners	47%
Following a template that has been successfully used by others	40%

Note: Regular usage includes responses of both 'often' and 'very often'.

Twenty percent of the respondents to this question made additional comments regarding other types of tactics used, including integrating current research and best practices from other fields (n=5), relying on particular instructional strategies (e.g., "problem-solving strategies, critical thinking, engaging activities, concrete experiences", etc.) (n=4),

emphasizing the input from key stakeholders (*n*=4), using certain strategies because they were mandated (*n*=4), trial and error (*n*=3), performance engineering (*n*=2), and repurposing materials (*n*=1) (Christensen & Osguthorpe, 2004). Of these, the authors were particularly interested in the use of instructional design and learning theories and they therefore asked respondents to list those theories that the used. Only 52% of subjects listed an instructional design theory and 50% of subjects listed a learning theory, although there was some overlap in responses to these questions.

Additionally, the study examined what types of sources respondents reported using in order to learn new theories, trends, and strategies. Respondents indicated regular use of (a) peer interaction (81%), (b) ID textbooks (51%), (c) websites (48%), (d) professional journals and magazines (48%), (e) literature from other fields (42%), (f) Education textbooks (33%), (g) professional conferences (28%), (h) Psychology textbooks (23%), and (i) Internet forums (19%).

With regard to practitioners who were generalists of performance improvement, Harless (1995) surveyed 23 organizations related to the preparation of practitioners in the field. A question relating to common sources of knowledge for the performance improvement professionals in their organization was included. Although intended as an indicator of professional preparedness, the findings also included more general sources of evidence as well. Both internal and external consultants (*n*=44) reported the two most common sources for skills within the organization: (a) coaching by other staff (*n*=14), (b) university coursework (*n*=10), (c) externally developed training (*n*=7), (d)

professional conferences (*n*=5), (e) textbooks and journal articles (*n*=4), (f) internally developed training (*n*=2), and trial and error (*n*=1).

Korth (1997) studied the planning processes used by professionals for planning a range of interventions and the underlying role of theory and pre-existing beliefs. These questions were explored within a small convenience sample of training and organizational development practitioners (*n*=5). Practitioners reported an iterative rather than a linear process of performance improvement. In terms of intervention selection, one subject reported consulting the opinions of others concerning a hypothetical solution—by "bounc[ing] it off of somebody" (p. 61). With regard to processes for identifying and generating alternatives, three subjects employed a stewing pot metaphor for intervention selection that reportedly relied upon intuition, dreaming, personal reflection, and flashes of genius. Despite subjects' views of simultaneously artistic and scientific approaches, a key implication of the study is that the process has moved from systematic, linear processes to a more holistic, chaotic, iterative, and creative design process.

The follow-on study performed by Korth (2000) examined the creative nature of performance improvement practices in more detail. In this study, the design process was characterized as having five phases: (a) diagnosis, (b) immersion, (c) percolation, (d) Aha! [breakthrough], and (e) checking. The first phase, diagnosis was identified as a rational process of developing a better understanding of the existing problem, circumstance, and desired outcome. Diagnosis also involved determination of a cause for the performance problem and a potential intervention to solve it. The second phase, immersion, was a phase where ideas were generated to specify further the characteristics

of an intervention. The third phase, percolation, involved an activation of the designer's subconscious, until a breakthrough occurred in the fourth phase. The fifth phase, checking, involved active processes to validate whether to move ahead with the idea generated in a breakthrough. This checking process could at first be informal, including activities such as confirming (a) their prior experiences, (b) fit with the desired outcomes, (c) relevance to content, (d) the experiences of others who have implemented similar interventions, (e) views of a respected colleague, and (f) getting the perspective of someone who has an opposing style.

Changes of Mind

Changes of mind may be defined as a desire and choice to switch from one option to another, based on Delaplace & Lescanne (2009). It is sparsely considered in the PI literature, typically as a mere implication or allusion. More commonly, there is a concern over perceived reluctance to give up prior beliefs in the face of scientific evidence:

> We get the impression that much of this distrust comes from a lack of support one finds in the research for people's intuition about the benefits of educational technology. Their reasoning seems to suggest that if research does not find evidence for something that seems so powerful, then research as an inquiry strategy must be flawed. (Clark & Estes, 1998, p. 5)

Here the implication is that practitioners hold tenaciously to prior beliefs to the point where they are willing to discredit counterevidence that is presented and are therefore unlikely to change their mind about how to proceed. Another example in the literature comes in the reintroduction to the field of Charles Pierce's ways of knowing (as cited in

Thomas, 2006, 2007). This framework includes reference to "tenacity" or "forming an opinion and stubbornly clinging to it" (p. 9).

There is general agreement that PI practitioners should defer making conclusions about what intervention(s) are most suitable until additional evidence is obtained. For the competency associated with how often "premature solutions offered by stakeholders" should be avoided, Guerra (2001b) and Guerra (2003) reported a median rating of five (i.e. *always*). An anecdotal example of this mindset is described by Hybert (2001):

> Any newcomer to the fields of training or performance technology is bound to notice the emphasis placed by these professions on analysis. Novices are often baffled by the number of different types of analysis—needs assessment, audience analysis, performance analysis, knowledge/skill analysis, goal analysis, meta analysis, etc. They are encouraged not to "jump to solutions" but to be sure and do their analysis first. Analysis gets a lot of press because it is, in fact, important and often done poorly or insufficiently. But analysis is only part of the picture. Through analysis you understand the problem, the situation, and the implications. Design involves making critical decisions about how to best address the problem, situation, and implications, given the complex set of stakeholder requirements, available resources, and environmental constraints that govern a specific project. The effective performance consultant needs to find ways to do sufficient analysis to make preliminary design decisions. Then additional focused analysis can be performed as needed for more detailed design decisions to eventually "spiral" to a

> solution that can then be implemented and produce the real end goal: performance. (p. 25)

Given the consensus of opinion regarding deferred intervention selection and the iterative nature intervention selection, it is somewhat surprising to find such little empirical consideration of changes of mind in the PI literature. After all, changes of mind would generate strong support for deferred intervention selection decisions. A second reason that the lack of consideration of changes of mind in the PI literature is unexpected is the great emphasis placed on changing the minds of clients. For example, Munley (2003) stated "When working with clients, either as an internal or external practitioner, requests for assistance often come in the guise of a request for training." (p. 18). Perhaps due to the field's roots in instruction, emphasis on changing clients' minds is particularly strong with regard to training interventions, but rather than calling for the death of training, Kaufman (2002) attacks means-based thinking instead, cautioning that "[t]raining is a means that can deliver useful results. However, before training can deliver its promise, before it can justify the time and money paid for it, we must first justify what we want as a result of training" (p. 5).

A notable example of empirical research relevant to changes of mind in performance improvement, Rowland (1992) found that despite the field's purported value of systematic solution selection, both experts and novices identified possible solutions early in their process. Both groups consulted the provided resource materials and their own memories, but novices maintained their initial selection and experts deferred commitment to their possible solutions and remained open to other possibilities;

furthermore, experts selected a variety of interventions but novices proceeded only with the instructional solution that was suggested in the problem materials.

Familiarity with Interventions

The experience of the PI professionals involved in a project is critical to its success (Swanson & Zuber, 1996) and it is widely accepted that PI practitioners ought to be well versed in a broad set of interventions (Hutchison et al., 1996; Hutchison & Stein, 1998; Medsker et al., 1995; Wellins & Rothwell, 2008). Even specialists within the field are expected to consider and be aware of a variety of interventions: advanced competency standards for instructional designers advocate that instructional designers must be able to consider and recommend non-instructional interventions when they are appropriate (Richey et al., 2001). Similarly, in addition to fundamental knowledge, entry-level human resource professionals are expected to possess knowledge of a variety of tactics that may be used enterprise-wide (Sincoff & Owen, 2004).

There is also an ethical obligation as a professional to consider a broad pool of interventions. Watkins, Leigh, and Kaufman (2000) pointed out that PI professionals should consider alternative solutions that are likely to resolve performance discrepancies, even if one's own organization does not offer or have expertise in these alternatives. Additionally, they suggested that practitioners are obligated to learn about new approaches, through reading journals outside the field, conference attendance, and colleaguial discourse.

From an academic perspective, PI educational programs appear to working to achieve this objective. Medsker and Fry (1992) reported a case study of a Master's level

human resource development program that embraced PI in its curriculum through both the incorporation of a results-oriented framework and an entire course devoted to non-instructional interventions. Subsequently, Medsker et al. (1995) surveyed 82 academic programs in performance improvement to determine the range of strategies covered. Their findings did not deal exclusively with intervention types (i.e., processes and techniques were also addressed). However, a ranked order of interventions with primary emphasis in the curricula may be extracted:

1. Training
2. Human resource management
3. Organizational design
4. Feedback systems
5. Strategic alignment
6. Personnel selection
7. Expert systems
8. Job aids and documentation
9. Job/work design
10. Performance support systems
11. Incentives
12. Ergonomics. (p. 15)

Despite the abundance of recommendations for practitioners' general familiarity with a variety of interventions, it is unclear to what extent the principle is applied in practice. What little relevant research that is available focuses on familiarity with and usage of

interventions as a function of experience. For example, Van Tiem (2004) surveyed 80 PI professionals on their self-reported familiarity with several types of interventions, including job analysis/work design, personal development, human resource development, organizational communication, organizational design and development, and financial systems. Here, a strong positive correlation between years in the field and solution expertise was observed (r=.508). The study also attempted to validate an expertise framework previously suggested by Hutchison, Stein, and Carleton (1996) and Hutchison and Stein (1998):

1. At an expert level, the practitioner has the ability to design a custom solution for any situation that can be defended (via evaluation) to expert specialists for 15 – 25 tactics across 10 or more areas.
2. Working competence involves the ability to design and implement 45 – 75 tactics across 15 or more areas.
3. Basic proficiency requires knowing the basic tenets and principles of half the tactics.
4. Partnership via contact with experts in all areas.

However, expertise takes several years to develop, as subjects with one to five years of experience had not attained the standards for expert or working competence. Yet, on average, the standard was met by years six to ten (Van Tiem, 2004).

Summary

This section reviewed the literature on the major variables involved in the present study. It showed that despite concern for the types and sources of evidence employed by

PI practitioners, there is a little empirical research on the topic. Studies tend to deal with instructional designers, who represent only a sub-set of performance improvement professionals and very few of these examine decision-making with a broad range of interventions. Therefore, the findings may not be generalizable to performance improvement professionals as a whole. In fact, given the sample sizes and convenience sampling methods utilized in a number of these studies (e.g., Mann, 1996; Rowland, 1992; Wedman & Tessmer, 1993a, 1993b; Winer et al., 1994), their findings may only be generalizable to *instructional designers* in a limited way. Furthermore, much of the research tends to emphasize procedures and deals with evidence usage only implicitly, generally requiring the reader to draw inferences about differences in usage of various types of evidence. Additionally, some concern has been raised regarding the tenacity with which practitioners hold their initial beliefs about suitable interventions. However the issue of changing one's mind appears to have been almost exclusively addressed in a conceptual, rather than empirical, manner. Finally, PI professionals ought to be familiar with a variety of interventions; yet, it is unclear to what extent this is the case in professional practice. One study that provided relevant data was Van Tiem (2004), although it did not address the role of familiarity in intervention selection.

All of these factors lend credibility that the study's research questions merit study. The next chapter addresses the methodological issues involved in providing rigorous answers to them.

CHAPTER 3 METHODS

The present study intends to answer general questions about changes in PI professionals' attitudes toward possible interventions during intervention selection and the roles of various types of evidence and self-reported familiarity with interventions in these changes. Within these general aims, the study focuses on the following research questions:

1. As they receive evidence, what changes occur in PI professionals' assessments of likely intervention success?
2. Which types of evidence do PI professionals find most persuasive?
3. Do PI professionals change their minds about what intervention is most likely to succeed?
4. Can PI professionals' changes of mind about what intervention is most likely to succeed be predicted by the nature of evidence received or assessments of likely intervention success?
5. When PI professionals change their minds about which intervention is most likely to succeed, are there differences in self-reported familiarity with interventions?
6. To what extent (if at all) are PI professionals' assessments of likely intervention success related with self-reported familiarity with interventions?
7. Do periods of practice in probabilistic reasoning influence professionals' assessments of likely intervention success?

This chapter details the methods that were used to answer these questions. The first part describes research design, variable specifications, the population and sampling

techniques. Next, the study's instrumentation will be discussed. This section describes the questionnaire that was used and the tactics employed in its development. Then, procedures for obtaining approval, delivering the questionnaire via the Web, the notification of subjects and the incentives for involvement and participation are outlined. The chapter closes with a description of the statistical analyses that were performed.

Research Design

This study employs a mixed (within- and between-subjects) design. The first research question used a 2x2x3 factorial design (scientific-craft nature x confirmatory nature x assessments of likely intervention success) with repeated measures on the third factor. A repeated measures design is especially appropriate given the Bayesian view of probability and the incremental confirmation principle discussed in previous chapters. On the other hand, within-subjects designs are not without issue as they may introduce a number of contextual effects, including practice, sensitization to differences in treatment, and carryover (Greenwald, 1976).

A variety of solutions have been proposed to minimize these concerns, including using a "wash out" phase, counterbalancing, and randomization (Crowder & Hand, 1990; Lamb, 2003; Ott & Longnecker, 2010; Runyon, Coleman, & Pittenger, 2000). The wash out approach introduces a gap of time to reduce the effects of the previous treatment (Namboodiri, 1972); therefore, it is not practically suited for a situation such as this. Counterbalancing can also be practically problematic because it can quickly lead to an unwieldy number of arrangements: where n equals the number of treatments, complete counterbalancing requires $n!$ order arrangements (Shuttleworth, 2009). Additional

complexity would be added in this case as the use of only one example from each type of evidence (i.e., scientific x agreement) would provide circumspect support for generalizations about PI professionals' reactions to each type of evidence. In a scenario where the number of arrangements would be linked to the number of items in the questionnaire and item analysis revealed a sufficient level of reliability with 12 items, the number of arrangements required for counterbalancing would total a practically infeasible 479,001,600. Therefore, the only remaining solution was to randomize questions.

Admittedly, randomization represents more of a "pseudosolution" to the possibility of introducing contextual effects such as practice, treatment sensitization, and carryover as "…it merely ensures that the contaminating effects are randomly distributed" (Pollatsek & Well, 1995, p. 790). However, randomization is a practice regularly employed to offset contextual issues in survey designs (Visser, Krosnick, & Lavrakas, 2000). Furthermore, analysis of the final research question dealing with the effects of practice on PI professionals' assessments of likely intervention success attends to this potential threat to validity. Techniques for executing this analysis will be discussed in the final section of this chapter.

Variable Specifications

This section operationalizes the variables targeted for study.

Likely intervention success. Subjects provided three consecutive probabilistic assessments (Pr_1, Pr_2, Pr_3) of an intervention's likely success on a verbal-numerical sliding scale from *0=(Almost) impossible* to *100=* (*Almost) certain*.

General assessment of likely success (Pr_1). The first probabilistic assessment of an intervention's likely success in general, on a verbal-numerical sliding scale from *0=(Almost) impossible* to *100=* (*Almost) certain*.

Prior probability (Pr_2). The second probabilistic assessment of an intervention's likely success on a verbal-numerical sliding scale from *0=(Almost) impossible* to *100=* (*Almost) certain*.

Posterior probability (Pr_3) The third probabilistic assessment of an intervention's likely success on a verbal-numerical sliding scale from *0=(Almost) impossible* to *100=* (*Almost) certain*.

Changes between posterior and prior probabilities (l). The natural log of a ratio of posterior probability to prior probability. This is a normalized measure of difference between prior and posterior probabilities.

Scientific nature of evidence. A dichotomous, categorical variable ("craft" versus "science"), where the categorization followed the classical continuum of science (Richmond, 1984) and ratings provided by an expert panel.

Evidential nature of agreement. A dichotomous, categorical variable (i.e. "infirming" versus "supportive" of one's initial intervention choice).

Initially-preferred intervention. Subjects indicated which type of intervention they believed to be most likely to resolve a gap in performance. Type of intervention varied in six ways, following Gilbert's Behavioral Engineering Model (1979).

Subsequently-preferred intervention. After receiving additional evidence, subjects indicated which type of intervention they believed to be most likely to resolve a gap in

performance. Again, type of intervention varied in six ways, following Gilbert's Behavioral Engineering Model (1979).

Changes of mind. Subjects' preferred intervention was converted into a dichotomous, dummy variable. The dummy variable had two levels: (1) sticking with the initially selected intervention and (2) switching to another intervention.

Self-reported familiarity. A semantic domain differential scale with five ordinal categories, ranging from *Not at all familiar* to *Highly familiar*.

Period of practice. Practice was treated based on the order in which the counterbalanced blocks of scenarios were presented ("I" was the first block the subject viewed, "II" was the second, and so forth).

Composite probabilities. The aforementioned probability assessments were individual measures provided by the research subjects. The first research question (RQ1) employed a mean, composite probability measure at each of the three probability assessments (Pr_1, Pr_2, Pr_3) across all scenarios, (i.e., $\Sigma(Pr_{1i})/N$, $\Sigma(Pr_{2i})/N$, $\Sigma(Pr_{3i})/N$ for each subject).

Population and Sampling

Performance improvement professionals work in settings including business, academia, government, health services, banking, and the military (ISPI, 2009) but most work in the consulting, finance, and educational service industries (Pershing, Cheng, & Foong, 2006). Generally, this study targets performance improvement professionals across all of these industries. However, this population was not easily accessible. When this is the case, the target population may be reconceived as an *accessible population*—

which is in itself a subset of the target population (Jones & Kottler, 2006). Therefore, the sampling frame was refined to certified performance technologists (CPTs).

The CPT designation is based on the Standards of Performance Technology and Code of Ethics first introduced in 2001 (Chevalier, 2008). The required performance standards include: (a) results-orientation, (b) systems-focus, (c) adding value, (d) partnership and collaboration, (e) systematic needs assessment, (f) systematic cause analysis, (g) systematic design, (h) systematic development, (i) systematic implementation, and (j) systematic evaluation (ISPI, 2000). Applicants for the CPT designation are required to show proficiency in these areas in three to seven projects (Hale, no date). The Code of Ethics includes principles of adding value, validated practice, collaboration, continuous improvement, integrity, and confidentiality. In addition to committing to uphold this code, CPTs must already have three years of work experience in the field and must agree to submit to recertification every three years.

The CPT designation is sponsored by the International Society for Performance Improvement (ISPI). ISPI members are located in 42 countries including the United States and Canada (ISPI, 2009). Performance improvement professionals also associate through other organizations such as the American Society for Training and Development or the Association for Educational Communications and Technology. However ISPI places a central focus on improving performance in the workplace—over particular interventions—and global membership (ISPI, 2009). Although sponsored by ISPI, CPTs need not be a member of this organization (Hale, no date).

Instrumentation

As previously mentioned, the study's research questions were addressed through the use of a questionnaire instrument delivered via the Web. Survey methods offer a variety of advantages, including efficient and economical collection of data that may be easily administered and analyzed (Creswell, 2003; Isaac & Michael, 1995; Marrelli, 2010; Patten, 2001). These characteristics complement the requirements and resources available for dissertation research. More importantly, the survey method can afford subjects relative anonymity (Isaac & Michael, 1995; Patten, 2001). This is especially important concerning sensitive matters, where the presence of an interviewer may influence the responses of a participant (Patten, 2001). This makes the survey method preferable to interview approaches in this case. More will be said about the merits of Web-based delivery in the Procedures section, as the added functionality allowed for investigation of the study's key variables.

There are a few potential disadvantages of using a questionnaire to collect data, including inability to confirm that respondents are the intended recipient of the questionnaire, that they understood the question, and low response (which may result in a non-representative sample) (Isaac & Michael, 1995). Mechanisms used to mitigate these risks are outlined in the study procedures.

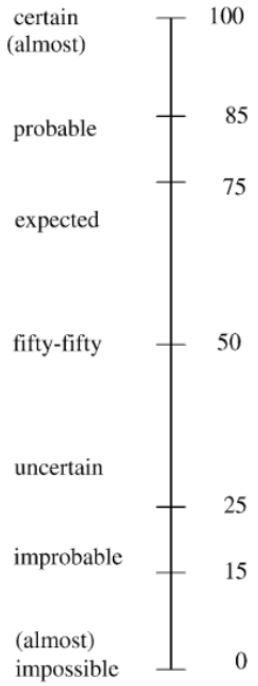
Questionnaire Description

The questionnaire consisted of four major sections: Welcome and Instructions, Background and Demographics, Experience, and Problem Scenarios. The Welcome and Instructions section provided a general description of the study's purpose and an

informed consent statement; subjects could not proceed to the questionnaire without acknowledging the informed consent statement. The Background and Demographics section asked subjects to indicate their gender, age in years, highest level of education that they have completed, the industry of their work organization, their work role (e.g., practitioner or researcher), and membership in other professional associations. Following this, the Experience section asked subjects to provide their experience in the field (in years), their familiarity with various types of interventions, and an initial probability assessment, Pr_1, in the form of general likely success. In the final section, Problem Scenarios, subjects responded to 12 scenarios by supplying several probability assessments (Pr_2 and Pr_3) and preferred intervention choices, separated by the receipt of additional evidence. A text-based version of the questionnaire is presented in Appendix B.

Likely intervention success assessments were elicited using a slider-response format. This response format is supported by the findings from a number of previous studies. Witteman and Renooij (2003) constructed and tested a combined verbal-numerical scale of probability with physicians, arts students, math students, and

information science students. The scale included seven verbal and numerical anchors, as



shown in Figure 4.

Figure 4. Combined verbal-numerical probability scale.

Test-retest reliability was established via correlational coefficients (r_s=0.752). Although discrete numbers and categories were provided, subjects were able to respond in a continuous manner; in fact, for the study of art and mathematics students, 80% of the responses were not directly tied to the anchors. This finding seems to support use of a continuous interval scale rather than ordinal ranges.

Additionally, the inclusion of verbal anchors seems to reduce the cognitive complexity of estimating probabilities: in the arts and mathematics subjects, Witteman and Renooij (2003) found no significant effects of the verbal-numerical scale on accuracy but found significant effects of the combined verbal-numerical scale on level of certainty. Furthermore, there were significant differences between groups who used the combined

scale and those who used a numerical scale only; those who used the combined scale found the problems they considered in the study easier, and appreciated the support of the scale more. Additionally, use of a continuous interval scale has some precedent in decision-making literature, (c.f. Chapman, 1973), who also made use of sliding number line technique. Chapman described the technique as follows:

> A 35-cm bar, displaying values from .50 to .99, was used for the subjects' probability estimates. One marker card, 7 ½ by 1 ½ cm, mounted on a 7 ½ -by-1 ½ cm base, was red; the other was blue. The subject chose either a red or a blue state of the world by selecting a marker card and made his probability estimate by placing the card at some value along the bar. (p. 272)

In this study, an electronic verbal-numerical slider response format was used, as illustrated in Figure 5.

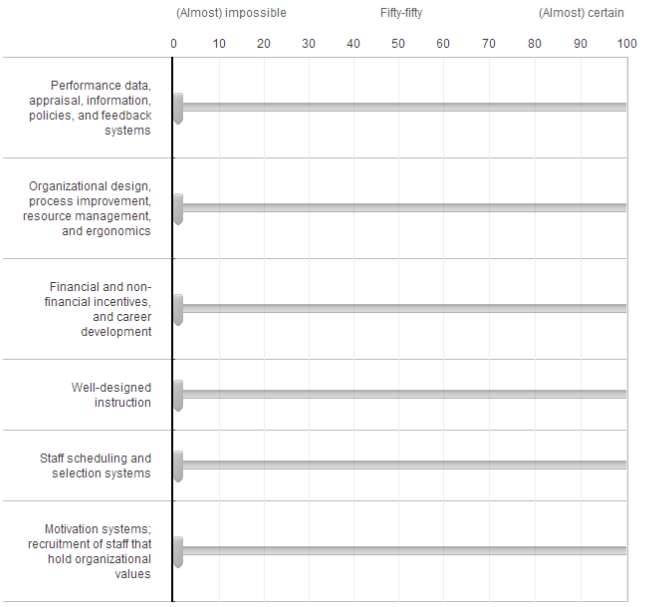


Figure 5. Slider response format for likely intervention success.

Validity and responsiveness

Preceding administration of the questionnaire, several activities took place to establish a valid and responsive tool. These activities are discussed chronologically.

Content and face validity. Problem scenarios were developed based on critical incident technique (Flanagan, 1954) and situational judgment test development procedures, following contemporary procedures reported by McDaniel & Nguyen (2001):

1. Collect critical incidents from subject matter experts.
2. Review and classify critical incidents.
3. Select representative incidents.
4. Edit incidents and develop situational item stems.
5. Assemble and administer a questionnaire to a second group of experts.
6. Have this second group identify possible courses of action for each situation.

The only deviation from these procedures is that step six, identification of possible courses of action, was not employed. This is due to the fact that subjects were asked to perform this task as part of the study.

Critical incidents (n=50) were collected from a convenience sample of PI professionals who are members of local ISPI chapters or who are part of the researcher's professional network. Appendix C includes the questionnaire that was administered in the face-to-face group session; a similar instrument was also fielded via a Web-based survey. For the face-to-face group sessions, an optional alternate activity packet was provided for those who were present but did not wish to participate.

Data were individually analyzed and then informal peer review discussions were also conducted as suggested by Creswell (2003) and Saldaña (2009). Each response was coded according to which interventions seemed likely, which were recommended, and which were implemented. Additionally, critical incidents were coded based on organizational background characteristics (i.e., industry) and the consultant's role in the organization (i.e., external or internal, or unstated).

Critical incident length ranged from five sentences to five pages; most were longer than is suitable for a survey instrument. Therefore, they were edited to form situational item stems, following the same general template: (a) name the organization and describe its industry or function, (b) specify the performance problem or opportunity, (c) express the client's concern or request, and (d) introduce further context and/or implications of the issue. Next, item stems were reviewed for consistent fidelity, complexity, clarity, and single-item structure (McDaniel, Morgeson, Finnegan, Campion,

& Braverman, 2001). Based on this review, a pool of possible problem scenarios (item stems) was identified (*n*=48).

Expert panels provide useful, evaluative, and critical input for content validation (Ary, Jacobs, & Razavieh, as cited in Guerra, 2001b; Tessmer, 1993). Therefore, a convenience sample of 30 performance improvement experts was contacted via email to participate in a Web-delivered questionnaire.

The response rate of this expert panel was 36.67% (*n*=11). The expert respondents' experience in the field ranged from 16 to 49 years, with a mean of 29.2. One respondent reported having a highest level of education as a Master's degree; all others noted having a doctoral degree. The experts also reported prolific publication in the field as well: the number of journal articles ranged from three to 275, *M=48, SD=78.48;* editorials ranged from 1 to 25, *M=9.75, SD=8.45;* book chapters ranged from one to 20, *M=7.56, SD=5.96;* and books ranged from 1 to 40, *M=9.80, SD=16.93*. Professional society involvement was broad-based as well, as shown in Table 5.

Table 5
Experts' Professional Society Membership

Professional Society	No. of respondents reporting membership
International Society for Performance Improvement (ISPI)	11
American Society for Training & Development (ASTD)	4
American Educational Research Association (AERA)	2
Association for Behavior Analysis International (ABAI)	2
Association for Educational Communications and Technology (AECT)	1
Society for Human Resource Management (SHRM)	1
Sigma Xi	1
American Society for Engineering Education (ASEE)	1
American Psychological Association (APA)	1
American Evaluation Association (AEA)	0

One expert also reported past membership in the American Education Curriculum and Development Association (AECD). Finally, most saw their work as varied, with 36.36% (*n*=4) identifying as being an "Author," "Practitioner," "Professor," and "Researcher;" only 18.18% (*n*=2) identified with only one of these roles. Experts most commonly identified with "Author" or "Practitioner" as descriptions of their work; 82% (*n*=9) identified with at least one of these roles. A smaller percentage, 55% (*n*=6) identified with "Professor" or "Researcher" as a description of their work.

Expert panelists were asked to rate the clarity and representativeness of each possible problem scenario on a 4-point semantic differential scale. As an attempt to meet requirements for face validity, experts also categorized various types of evidence. A pool of general types of evidence was based on review of the literature (*n*=125). Panelists rated each type of evidence as either artistic or scientific, then on a seven-point semantic

differential scale rated the extent to which this type of evidence exemplified the selected category. Anchors were presented for each point, ranging from *Highly craft-based*, on the low end to *Highly scientific inquiry*, at the high end.

For the expert panel, responses to questions regarding both the performance improvement scenarios and to the evidential item pool were analyzed using the Fleiss (1971) kappa statistic (κ). A fair level of agreement amongst raters existed (κ =0.2887267*,* $p < .001$*)* (Landis & Koch, 1977). Given that only a fair amount of agreement was indicated and that Fleiss' kappa can underestimate agreement of ordinal data (Hripcsak & Heitjan, 2002), the intraclass correlation coefficient (*ICC)* was also calculated as a measure of inter-rater reliability. This analysis focused on scenario representativeness, as a low level of reliability on this variable might have suggested that insufficient face validity existed for the tested scenarios (perhaps justifying development of additional scenarios) However, in a two-way mixed model with measures for consistency *ICC*=.614, 95% CI [.331,.811].

Following this phase, an initial questionnaire with 16 scenario items was developed. Scenarios were selected based on an ordered ranking of the percentage of experts who marked the highest levels of representativeness and clarity, as demonstrated in Appendix F.

Types of evidence were selected based on an ordered ranking of the ratio of experts who marked a type of evidence as science-based to those who marked it craft-based, as illustrated in Appendix G.

Evidence descriptions were reviewed for clarity and consistency. A key element of consistency was description length, in number of words. At first, description lengths were non-normal (M=92.875, SD=29.81471, s^3= .518518, s^4= -.80318); however, several rounds of editing resulted in a less skewed, but somewhat platykurtic distribution (M=92.4375, SD=24.38844, s^3= 0.145602, s^4= -1.24137). There were no differences in length between artistic and scientific evidence descriptions (U=18, p = 0.15595). The reading level of evidence descriptions were high, according to the Flesch-Kincaid grade level statistic (M=12.375, SD=3.048481, s^3= 0.063272, s^4= 0.43537); no differences in grade level were exhibited between artistic and scientific evidence descriptions (U=43, p = 0.27863). A smaller expert panel provided informal feedback on the updated questionnaire.

The resulting questionnaire was piloted with a sample of CPT's (n=52). Table 6 shows that gender mix was approximately equal:

Table 6

Pilot Sample Characteristics – Gender Mix

Gender	No.	%
Female	27	51.92%
Male	25	48.07%

Subjects' ages ranged from 26 to 85 years (M=56.42, SD=10.68). In fact, almost three-quarters of the subjects were age 50 and over (n=38, 73.07%).

Most subjects reported having a Master's (61.538%) or Doctoral degree (25.00%). A small number indicated a Bachelor's (n=3) or a high school degree (n=1). Three subjects noted 'Other' and explained that their highest level of education was a

technical diploma, certificate program, or being a Doctoral candidate. No subjects held an Associate's degree as their highest level of education.

With regard to their work organization's industry, 11 subjects elected to mark "Other" rather than selecting one of the general sectors included in the North American Industry Classification System (NAICS); however, more detailed descriptions of the sectors are available (US Department of Commerce, 2012). These detailed descriptions were used to recode 10 responses. One subject indicated "Other," but did not provide any text to explain the subject's organization's industry. Another subject did not respond to this item. As shown in Table 7, most subjects worked in Professional, Scientific, Technical Services, and Consulting; Educational Services; Health Care and Social Assistance; or Public Administration.

Table 7

Pilot Sample Characteristics — Work Organization Industry

Industry Sector	No.	%
Professional, Scientific, Technical Services, and Consulting	18	36.00%
Educational Services	9	18.00%
Manufacturing	6	12.00%
Health Care and Social Assistance	5	10.00%
Public Administration	4	8.00%
Information	2	4.00%
Utilities	2	4.00%
Finance and Insurance	1	2.00%
Management of Companies and Enterprises	1	2.00%
Mining, Quarrying, and Oil and Gas Extraction	1	2.00%
Retail Trade	1	2.00%

Only a small percentage (5.769%) classified themselves only in the role of a "Researcher." Most characterized themselves in the role of a "Practitioner" (67.308%),

while a moderate group (26.923%) identified with both the "Researcher" and "Practitioner" roles.

Of the professional societies offered to subjects, most subjects participated in ISPI (n=38). Four subjects were members of ASTD. Very few did not answer (n=3) and only one subject noted not belonging to any professional associations. On the other hand, six subjects conveyed that they were members of multiple professional associations, while ten subjects listed professional associations not provided in the options (e.g., Project Management Institute, International Federation of Training and Development Organisations).

In terms of years in the field, the experience level was quite high (M=23.630, SD=10.673).

A key aim of the pilot was to test whether changes occurred between prior probabilities and posterior probabilities. Therefore, a log-likelihood ratio across subjects was calculated for each scenario. Here, log likelihood (l) values of zero indicate no changes between prior probabilities (Pr_2) and posterior probabilities (Pr_3). When $l > 0$, changes favor posterior probabilities; when $l < 0$, prior probabilities are higher. As illustrated in Appendix F, the direction of changes between prior probabilities (Pr_2) and posterior probabilities (Pr_3) corresponded with the agreeable nature of evidence provided in each scenario: positive values were only associated with supportive evidence and negative values were only linked with non-supportive evidence. This observation supports the face validity of the instrument.

Responsiveness. Instrument reliability is often measured using Cronbach's alpha (α), but this is suitable only for summative scales (Peterson, 1994; Santos, 1999). Alternatively, evaluative scales should be assessed for their sensitivity to change, which is also known as "responsiveness" (Terwee, 2003). Although consensus about which measure of responsiveness performs best, Husted et al. (1999) noted that when internal responsiveness is of concern, Chi square models provide a suitable and interpretable measure of responsiveness. As such, the l measure for each scenario (as reported in Appendix F) was converted to $2l$, which can be shown to follow closely the Chi-square distribution (Ghosh, personal communication, October 5, 2013) and then tested for responsiveness.

A second aim of the pilot test was to reduce the length of the instrument. Of note, this aim was supported in the qualitative responses to the final question in the pilot, which invited subjects to provide comments about their assessments or the questionnaire in general. Of the 22 subjects that offered comments, 27% (n=6) remarked on the length of the questionnaire and time required to complete it.

Therefore, the log likelihood and responsiveness measures within each group of scenarios were reviewed to determine which scenarios could be eliminated. Scenarios were eliminated if they were not shown to be responsive or if the log likelihood measure was atypical for the group. Appendix F summarizes the results of this analysis.

Procedures

Various procedures were used to garner access to the sample population, deliver the questionnaire instrument via the Web, notify potential subjects about the study,

stimulate involvement and participation, obtain resources, and to protect human subjects from harm. The related procedures for each of these aspects will now be discussed in greater detail.

Approval

In order to obtain permission to execute the study and to attend to the practical matter of gaining access to email addresses for the population, ISPI's Executive Director was contacted by phone and approval was confirmed by email (Davis, personal communication, October 6, 2010).

Web-based delivery

Hoonakker and Carayon (2009) distinguished between various types of Internet-based questionnaires. In their view, Internet surveys may be embedded in an email message, attached to an email as a document, attached as an executable file, or placed and stored on a Web server. The present study employs the latter method and is therefore a Web-based questionnaire. Web-based delivery is especially suitable in this case due to the conditional structure of the evidential confirmatory nature variable which necessitates the "piping" of questions and answer text based on subjects' previous responses. Admittedly, a similar approach could be employed by an interviewer. However Web-based delivery of the questionnaire handles the issue much more efficiently and with less likelihood of error.

Of note, some concern has been raised about the effects of Web-based delivery of surveys upon response rates (Hoonakker & Carayon, 2009; Kaplowitz, Hadlok, & Levine, 2004). In a previous meta-analysis of Web and Internet-based surveys, Cook,

Heath, and Thompson (2000) found that the number of notifications and incentives may be effective in increasing response rates; therefore both tactics were applied.

Notifications

The sample population was contacted through a series of three emails (at the address that is provided from ISPI's membership database). The first email contact, Invitation Email, (included in Appendix G) discussed the purpose and importance of the study, estimated time required for completion, a link to the electronic survey, information about the incentive that was being offered for study participation, and the researcher's contact information (for use in case technical issues were encountered during completion of the questionnaire). Two subsequent reminders were sent. The Reminder Email is also provided in Appendix G.

Incentives

As a means for encouraging participation and reducing the risk of response bias, a contingent incentive was utilized. As opposed to noncontingent incentives, which are provided to the entire sample, the provision of contingent incentives is dependent upon completion of the survey task (Trussell, 2008). The incentive for the study was a $25 certificate from www.restaurant.com. Incentive recipients were sent email notices including a link for redeeming their gift certificate.

Statistical Analysis

A major consideration for the design and execution of this study was the selection of statistical tests to analyze its research questions. Due to the basis of the study in Bayesian principles, this decision was especially complex. For example, if frequentist

approaches were utilized, then the underlying theory of probability employed within them would run counter to the primary questions targeted by the study. At the same time, as measured by the number of articles that employ them, the application of Bayesian methods in education, social sciences, economics and econometrics, law, quality management, and medicine has grown dramatically since the 1970's—roughly doubling every decade (Berger, 2000). However, they are not well known within the field of performance improvement (Pershing, 2008). As limitations were present for both approaches, the research questions were analyzed in both traditions where possible.

The first research question (RQ1) dealt with changes in PI professionals' assessments of likely intervention success between different types of evidence, namely categorical levels of scientific and agreeable nature. As explained earlier, there were repeated measures on the third factor. RQ1 was analyzed with a 2x2x3 repeated measures ANOVA. The second research question, (RQ2) was concerned with the types of evidence that PI professionals view to be persuasive. Because there are two dependent variables (i.e., the level of the posterior probability assessment and the compound measure of the differences between this posterior probability assessment and prior probability assessments), this relationship was analyzed using a 2x2 Factorial MANOVA. MANOVA is not robust to violations of assumptions of independence, so a separate test was conducted for each of the three, randomized blocks of scenarios. The third research question (RQ3) addressed whether PI professionals changed their minds about which intervention would be most likely to succeed and was tested via Normal (*Z)* approximation. The fourth research question (RQ4) was concerned with the prediction of

changes of mind. Three separate analyses were conducted. The first examined the scientific nature of evidence and changes of mind, while the second tested the evidential nature of evidence and changes of mind. Both were analyzed using Spearman's r_s. The final analysis in RQ4 was a binary logistic regression of changes between posterior and prior probability, l; the level of posterior probability, Pr_3; and changes of mind. The fifth research question (RQ5) explored differences between cases where PI professionals changed their minds and where they did not, with regard to self-reported familiarity with either an initially-preferred or subsequently-preferred interventions. Two separate Mann-Whitney U tests were used to analyze this question. RQ6 asked about the relationship between the continuous interval assessments of likely intervention success and self-reported familiarity with the selected intervention. In this analysis, the self-reported familiarity for the selected intervention type was considered and it was tested for correlation with the corresponding probability assessment (i.e., Pr_1, Pr_2, Pr_3) using separate executions of Spearman's r_s. The final research question (RQ7) was concerned with the effects of practice on PI professionals' assessments of likely intervention success. This question was studied with repeated measures ANOVA.

CHAPTER 4 RESULTS

Before more detailed findings are presented, descriptive statistics on the study's sample are reviewed.

Sample Characteristics and Response

There were 701 active CPTs at the time of the study's administration (i.e., January 9 – February 6, 2014). Of these, 680 had a certification date recorded in the database. These certification dates ranged from early 2002 to late 2013; mean certification length was 7.918 years (*SD*=0.14, approximately 51 days). American residents represented 88% of the sample frame (*N*=617), while the remaining 22% were from other countries of origin. Estimating on the basis of first and/or middle names, almost half were females (49.50%) and slightly fewer CPTs were male (44.94%), with the caveat being that 5.56% (*n*=39) bore a name that was indeterminate or gender neutral according to conventional wisdom.

Some active CPTs (*n*=28) were excluded due to a missing email address. In total, 673 were contacted via email to solicit participation in the study. As demonstrated in Figure 6, there were high rates of non-participation and non-completion.

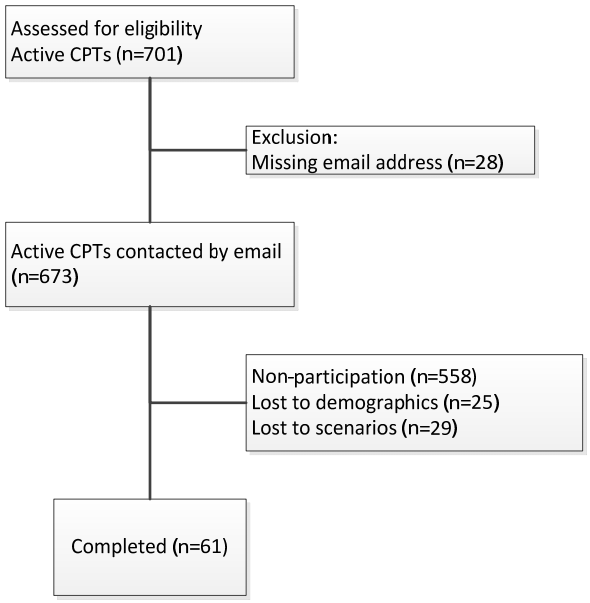


Figure 6. Subject flow diagram.

Of note, some of the non-participation and non-completion may be attributable to tactics employed to maintain independence between the pilot and final administrations. This came in the form of a request in the email for subjects in previous phases of the study not to participate in further data collection. At the time of the questionnaire's close, 61 responses were completed; this equated to a 9.06% response rate.

A variety of demographic variables were collected from subjects at the beginning of the questionnaire: gender, age, highest level of education, work organization industry, work role, professional association membership, and years of PI experience. Characteristics of the sample are discussed in turn.

Gender mix was approximately equal, as summarized in Table 8.

Table 8.

Sample Characteristics - Gender Mix

Gender	No.	%
Female	32	52.459%
Male	29	47.540%

Subjects' ages ranged from 33 to 81 years (*M*=56.98*, SD*=9.75*)*. More than three-quarters of the subjects were age 50 and over (*n*=46, 76.666%). One respondent did not enter an age.

Most subjects reported having a Master's degree (55.74%) or Doctoral degree (24.59%). Those who indicated "Other" had completed a post-graduate certificate program or Doctoral coursework (without completion of a dissertation).

Table 9

Sample Characteristics - Education Level

Highest level completed	*N*	%
Associate's degree or certificate	4	6.56%
Bachelor's degree	5	8.20%
Doctoral degree	15	24.59%
Master's degree	34	55.74%
Other (please list)	2	3.28%
No answer	1	1.64%

With regard to their work organization's industry, 14 subjects elected to mark "Other." Thirteen of these responses were recoded similar to the pilot study, using the NAICS classifications (US Department of Commerce, 2012). One subject indicated "Other," but did not provide any text to explain the subject's organization's industry. As

shown in Table 10, most subjects worked in Professional, Scientific, Technical Services, and Consulting; Educational Services; or Public Administration.

Table 10

Sample Characteristics - Work Organization Industry Rankings

Industry Sector	*f*	%
Professional, Scientific, Technical Services, and Consulting	24	39.34%
Educational Services	9	14.75%
Public Administration	5	8.20%
Manufacturing	4	6.56%
Management of Companies and Enterprises	3	4.92%
Mining, Quarrying, and Oil and Gas Extraction	3	4.92%
Retail Trade	3	4.92%
Utilities	3	4.92%
Finance and Insurance	2	3.28%
Health Care and Social Assistance	2	3.28%
Information	2	3.28%
Other	1	1.64%

No subjects classified themselves only in the role of a "Researcher," but approximately a fifth of the subjects identified with both the "Researcher" and "Practitioner" roles (*n*=12). By and far, most subjects characterized themselves as a "Practitioner" exclusively (*n*=49, 80.327%).

Of the professional societies offered to subjects, most subjects participated in ISPI (*n*=52) or the American Society for Training & Development (*n*=9). Several subjects conveyed that they were members of multiple professional associations (*n*=9), while only a few listed professional associations not provided in the options (e.g., American Society for Quality, American Nuclear Society, eLearning Guild). Three subjects did not indicate membership in any professional association.

Again, the experience level was quite high *(n*=56, *M*=22.589, *SD*=10.990). Four subjects did not indicate their years of experience in the field and one was excluded because a negative integer was entered.

To get a sense of the bias that may have been introduced by response, subjects with partial responses were compared to the study sample on key demographic variables. These groups were similar on gender, (X^2=0.1306, *p*= 0.717759), age (*U*=1290.5, *p*=.38667), years of experience in the field (*U*=1014, *p*=.78567). Comparing probability measures yielded no differences either: Pr_1 (*U*=34116, *p*=.129698), Pr_2 (*U*=23752, *p*=.077416), or Pr_3 (*U*=19050, *p*=.075122).

The next section reports the findings of the statistical analyses that were performed on the study's research questions. Each question is addressed in turn. Where both frequentist and Bayesian tests were conducted, frequentist results are reported first, followed by the Bayesian results. Frequentist tests were performed using the SAS 9.3 software package and Bayesian statistics were calculated using WinBUGS version 1.4 with an *R*-language interface. An α level of .05 was used for all frequentist statistical tests; Bayesian tests required that the 95% Credible Interval did not encompass the zero value (Ghosh, personal communication, May 25, 2014). Mean estimates from the pilot study were used as informative priors for the Bayesian tests.

Changes in Assessments of Likely Intervention Success

RQ1 focused on what changes occur in PI professionals' assessments of likely intervention success as they received evidence. A score for each assessment and each 2x2

factor combination was calculated for each subject: $\Sigma(Pr_{1i})/N$, $\Sigma(Pr_{2i})/N$, $\Sigma(Pr_{3i})/N$ Means and standard deviations are provided in Table 11.

Table 11

Descriptive Statistics for Composite Assessment of Likely Intervention Success by Nature of Evidence

	Pr_1		Pr_2		Pr_3	
	M	*SD*	*M*	*SD*	*M*	*SD*
Scientific, Supportive	66.95	16.82	72.15	12.69	80.33	12.35
Craft, Supportive	67.45	17.59	67.16	15.79	80.82	13.36
Scientific, Infirming	68.24	16.25	70.86	11.95	37.95	20.29
Craft, Infirming	69.73	16.45	69.44	12.71	40.83	20.22

Figure 7 shows mean composites by nature of evidence at Pr_1, Pr_2, and Pr_3.

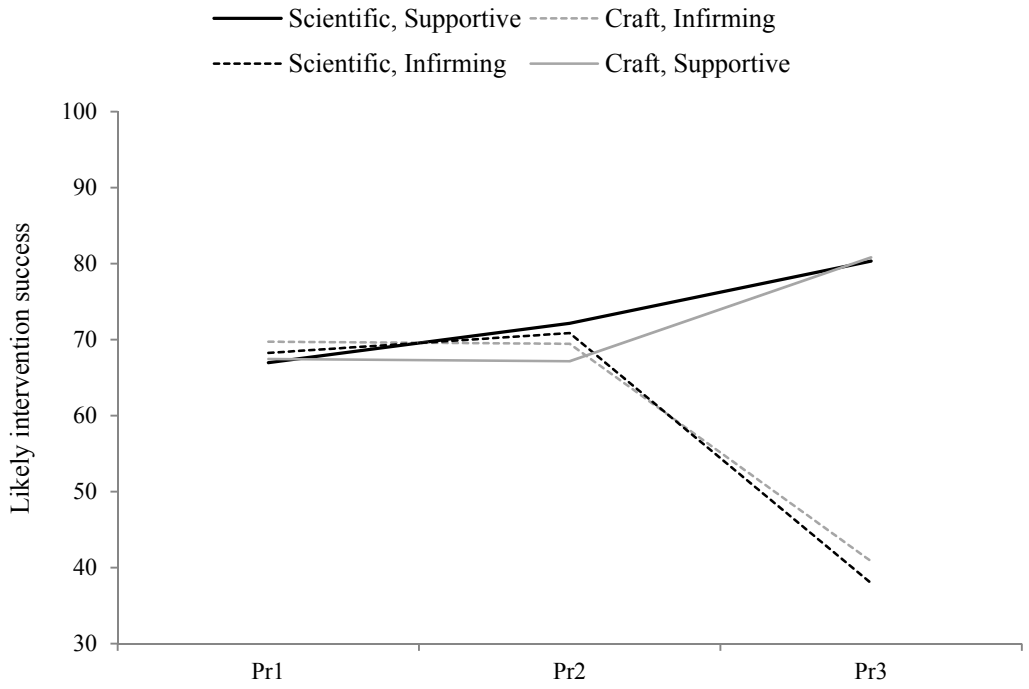


Figure 7. Mean composite assessments of likely intervention success by nature of evidence.

A 2x2x3 repeated measures analysis yielded main effects for time, F (1, 665) = 6.51, p<.001; and nature of evidential agreement, F (1, 665) = 28.83, p<.001. More

importantly, an interaction effect between time and nature of agreement was observed F (1, 665) = -21.48, p<.001. No effects were noted for scientific nature of evidence.

Bayesian analyses were also performed for RQ1, with mixed results. Table 12 reports posterior estimates.

Table 12

Posterior Estimates for Assessments of Likely Intervention Success – Bayesian

Parameter	Mean	*SD*	95% Approximate Credible Interval	
Scientific	3.075	2.340	-1.5670	7.680
Supportive*	10.823	2.398	6.1112	15.544
Time*	-3.593	1.064	-5.6515	-1.522
Time x Scientific	-1.211	1.209	-3.5714	1.107
Time x Supportive	2.073	1.231	-0.3393	4.495

Here, main effects for evidential agreement and time were confirmed, but no other effects were. More will be said about this in the Discussion section of Chapter 5.

RQ2 was concerned with which of these types of evidence were found to be most persuasive by PI professionals as a function of changes between prior and posterior probabilities (l) and the level of posterior assessment of likely intervention success (Pr_3). It was studied using three separate Dunnett-Hsu corrected 2x2 factorial MANOVA tests (one for each block of scenarios). In all three analyses, nature of evidential agreement was significant: F(2, 221) =94.45, Wilks' Λ= 0.54, p <.001; F(2, 224) =111.10, Wilks' Λ= 0.50, p <.001; and F(2, 228) =96.04, Wilks' Λ= 0.54, p <.001. Scientific nature of evidence was only significant in Block B: F(2, 224) =7.54, Wilks' Λ= 0.94, p =.0007. Follow-up univariate analyses within that same block revealed no significant differences

on the level of posterior assessment of likely intervention success (Pr_3); however, significant differences were noted for the change between prior and posterior probability measure (l), p=.0006.

Changes of Mind

The next set of research questions dealt with changes of mind about what intervention is most likely to succeed. RQ3 was concerned with whether PI professionals changed their minds about what intervention would succeed as they received evidence. Changes of mind were treated discretely via a dummy variable: (1) sticking with the initially selected intervention and (2) switching to another intervention. The question was tested using Normal (Z) approximation, where:

H_0: p =.50

H_1: $p \neq .50$

Across all of the scenarios, subjects stuck with their initial intervention choice in 472 cases and switched to another intervention in 256; therefore, the observed proportion is .6484. The frequencies of sticking with an initial choice and changing one's mind were significantly different (Z=8.0055, p <.0001). Individual scenarios were also analyzed; only one case, AD5, did not have significant differences between instances of sticking and switching interventions (observed proportion of sticking=.5574, Z=0.8963, p=0.3701).

RQ4 was concerned with forecasting changes of mind. The first analysis looked at associations between the type of evidence received and changes of mind about which intervention was preferred, via two separate Spearman's r_s tests. In the frequentist

analyses, no significant correlation was established between Nature of agreement and changes of mind (r_s =-0.0345, p=0.1757). However, the Bayesian tests illustrated a strong association between Supportive evidence and sticking to an initial intervention (r_s=-0.5509, α=.05, HPD Interval =-0.283, -0.0834). Scientific evidence was moderately correlated with sticking to an initial intervention choice (r_s =-0.3160, p<.0001). Bayesian tests noted a significant but almost negligible correlation with switching (r_s=.0283, α=.05, HPD Interval =-0.0832, -0.0572).

The third analysis in RQ4 examined if changes of mind could be predicted by the assessed levels of likely intervention success, looking at both posterior probability (Pr_3) and changes between prior and posterior probabilities *(l)*. Again, changes of mind were treated discretely. Forty-seven cases were deleted due to missing observations on either of the variables. Of the remaining cases, there were 468 instances where subjects stuck with their initial intervention choice and 217 occurrences where subjects changed their mind about which intervention they preferred.

Posterior probability and changes between prior and posterior probabilities are good models for changes of mind: all three Chi-Square tests for goodness of fit were significant p<.0001 and Max-rescaled R^2=.75. Table 13 illustrates that both Pr_3 and l were significant; Table 14 shows Odds Ratio estimates.

Table 13

Maximum Likelihood Estimates for Changes of Mind - Frequentist

Parameter	*df*	Estimate	*SE*	Wald Chi-Square	*p*
Intercept	1	-3.3767	0.7	23.269	<.0001
Pr_3	1	0.082	0.0112	53.9489	<.0001
l	1	1.3329	0.4854	7.5394	0.006

Table 14

Odds Ratio Estimates for Changes of Mind – Frequentist

Effect	*OR*	95% Wald Confidence Limits	
Pr_3	1.085	1.062	1.109
l	3.792	1.464	9.820

Tables 15 and 16 report analogous results for the Bayesian tests, while Figure 8 illustrates density plots.

Table 15

Posterior Estimates for Changes of Mind – Bayesian

Parameter	*df*	Mean	*SD*	95% Approximate Credible Interval	
Intercept	1	-3.4225	0.73036	-4.92375	-2.0512
Pr_3	1	0.0832	0.01163	0.06173	0.1071
l	1	1.3534	0.49753	0.38425	2.3596

Table 16

Posterior Odds Ratio Estimates for Changes of Mind – Bayesian

Effect	*OR*	95% Approximate Credible Interval	
Pr_3	1.086	1.064	1.113
l	3.871	1.468	10.587

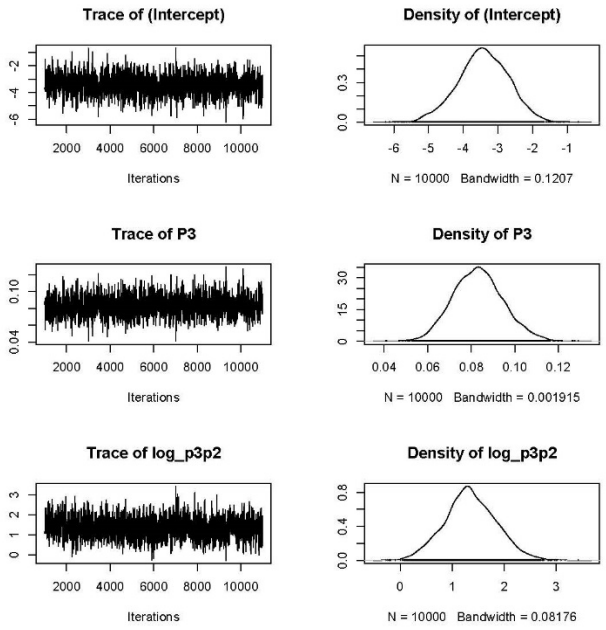


Figure 8. Density plots for changes of mind.

RQ5 examined whether there were differences in self-reported familiarity with interventions when changes of mind occurred. When changes of mind occurred, there were no differences in self-reported familiarity on initially-preferred interventions (Z=-1.7215, p =.0852), but differences were observed for subsequently-preferred interventions (Z=-2.9722, p = .0030). Ad hoc paired t-tests were also conducted to compare self-reported familiarity on selected interventions with a composite measure of self-reported familiarity on the other non-selected interventions. Again, these tests looked

at both the initially and subsequently preferred interventions. In both cases, levels of self-reported familiarity were higher for the selected interventions than their non-selected counterparts, $t(730) = 9.27$, $p<.0001$; $t(729) = 6.57$ $p <.0001$. The next section reports more detailed analytical results on self-reported familiarity.

Self-reported Familiarity with Interventions

High levels of familiarity with all interventions were reported in the sample, as Figure 9 illustrates. Across all intervention types, 65.02% of ratings were either *highly* or *very familiar.*

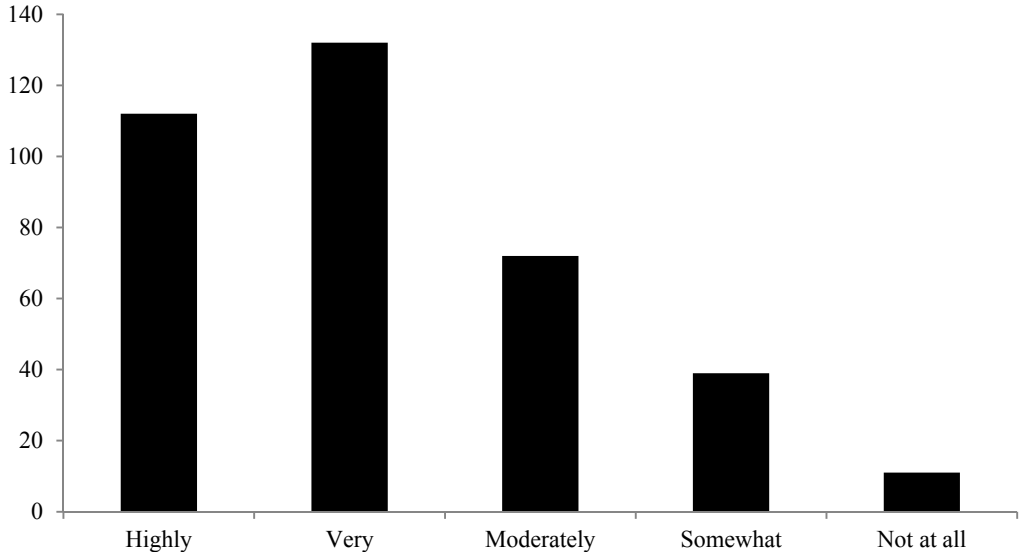


Figure 9. Familiarity ratings across all intervention types.

A negligible positive correlation between years of experience and self-reported familiarity was observed (r_s=0.0563).

Knowledge interventions showed the highest levels of self-reported familiarity: 62.3% of subjects indicated that they were "highly familiar'" with "Well-designed

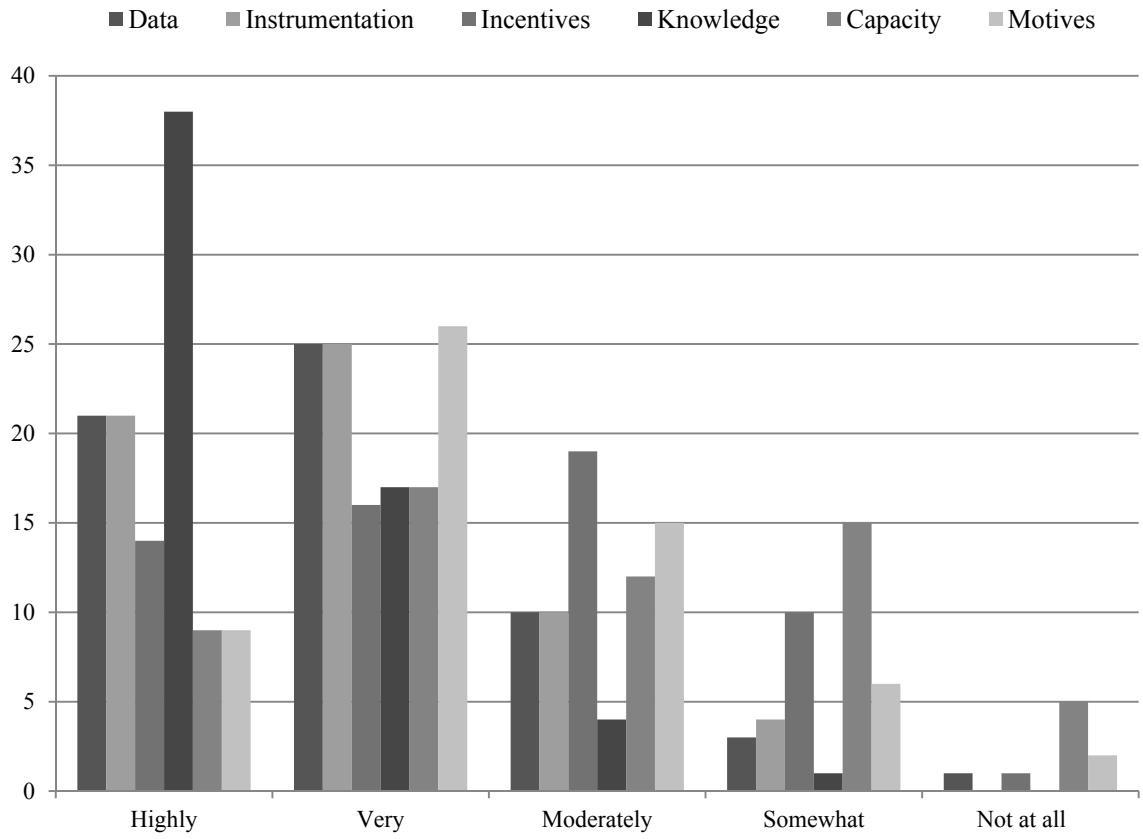


Figure 10. Familiarity ratings by intervention type.

instruction." As shown in Figure 10, the lowest levels of familiarity were reported for Capacity interventions, (i.e. "Staff scheduling and selection systems.") and Motives interventions (i.e. motivation systems; recruitment of staff that hold organizational values).

In the present sample, eighteen separate Spearman's r_s tests were conducted correlating self-reported familiarity with assessments of likely intervention success (i.e., Pr_1, Pr_2, Pr_3). As illustrated in Table 17, few significant correlations existed for RQ6. For general assessments of likely intervention success (Pr_1), strong positive correlations were present for self-reported familiarity of Knowledge and Motive interventions (r_s=0.43 and

0.53, respectively). More moderate positive relationships were observed for prior assessments of likely interventions success (Pr_2) and Incentives (r_s=0.38) and Knowledge (r_s=0.206). Of note, no significant correlations occurred between self-reported familiarity and posterior assessments of intervention success (Pr_3).

Table 17

Familiarity and Likely Intervention Success Correlations - Frequentist

	Pr_1			Pr_2			Pr_3		
	r_s	n	p	r_s	n	p	r_s	n	p
Data	0.132	195	.065	0.14	195	.0504	0.044	188	.5439
Instrumentation	-0.006	245	.917	0.09	245	.149	0.005	234	.9362
Incentives	0.169	42	.283	0.38	42	.012*	0.239	40	.1375
Knowledge	0.430	150	<.0001*	0.206	150	.011*	0.162	145	.0504
Capacity	-0.231	44	.130	-0.02	44	.8859	0.167	42	.2879
Motives	0.530	55	<.0001*	0.119	55	.385	-0.14	53	.3093

When results of the pilot sample were included as priors (see Appendix H) for the Bayesian analyses, results differed. As shown in Table 18, only three comparisons did not illustrate significant correlations: Instrumentation (Pr_1), Capacity (Pr_2), and Instrumentation (Pr_3).

Table 18

Familiarity and Likely Intervention Success Correlations - Bayesian

	Pr_1			Pr_2			Pr_3		
	α	*HPD* Interval		α	*HPD* Interval		α	*HPD* Interval	
Data	0.1372*	0.1100	0.1643	0.1452*	0.1180	0.1721	0.0497*	0.0220	0.0771
Instrumentation	-0.00139	-0.0291	0.0261	0.0975*	0.0700	0.1248	0.0105	-0.0172	0.0380
Incentives	0.1724*	0.1453	0.1993	0.3829*	0.3588	0.4069	0.2412*	0.2148	0.2674
Knowledge	0.4336*	0.4108	0.4561	0.2108*	0.1842	0.2372	0.1674*	0.1404	0.1942
Capacity	-0.2238*	-0.2503	-0.1976	-0.0167	-0.0445	0.0109	0.1711*	0.1439	0.1980
Motives	0.5311*	0.5106	0.5514	0.1236*	0.0962	0.1508	-0.1357*	-0.1630	-0.1087

Density plots are included in Appendix I.

Practice in Probabilistic Reasoning

Finally, RQ7 explored how the role of practice may influence professionals' assessments of likely intervention success. Here, Pr_2 and Pr_3 were examined in terms of the randomized order in which scenarios were presented: Block I, Block II, and Block III. Descriptive statistics are reported in Table 19.

Table 19

Likely intervention success descriptive statistics by blocks

	Pr_2		Pr_3	
	M	*SD*	*M*	*SD*
Block I	70.52869	12.07935	60.1694	13.32832
Block II	68.81148	13.17302	58.59973	13.97574
Block III	70.43033	14.81181	61.59973	16.43825

Repeated measures ANOVA showed no significant interaction effect, F (2, 300) = 0.22, p=0.806. In fact, Blocks I, II, and III were similar (M=65.35, 63.71, and 66.02, p<.0001) and an ANOVA revealed no significant differences between Block I and II (p=.201), Block I and III (p=.604), or Block II and III (p=.072). Bayesian repeated measures ANOVA also showed no significant interaction between time and order (M=0.6872, SD=0.8055, 95% CI, -0.8995, 2.26217). Figure 11 presents density plots for this analysis.

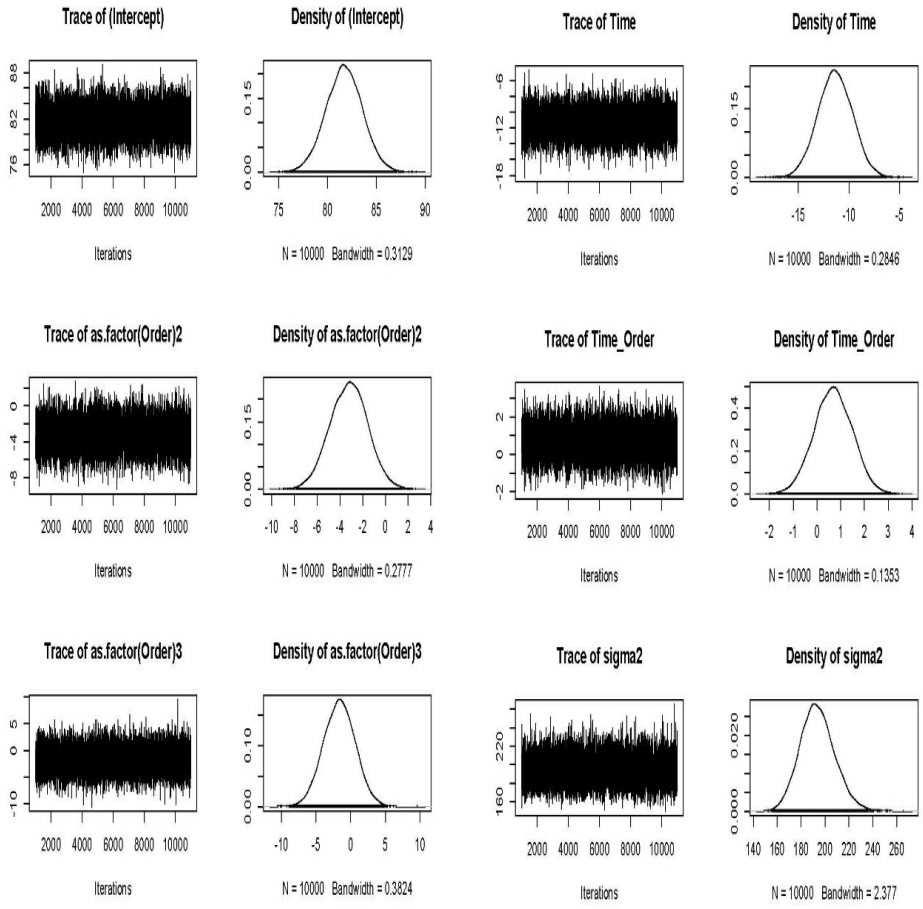


Figure 11. Density plots for practice in probabilistic reasoning.

CHAPTER 5 DISCUSSION

The purpose of this study was to contribute empirical knowledge about PI professionals' intervention selection decisions, with special attention to changes in assessments of likely intervention success and changes of mind about preferred interventions during this process. The roles of self-reported familiarity with interventions and practice in probabilistic reasoning were also examined. These aims were executed via a dynamic, Web-delivered questionnaire. Where feasible, statistical analyses were conducted in both frequentist and Bayesian paradigms. Mean estimates from the pilot test were used as informed priors for the Bayesian tests and in some cases mixed results were noted.

Evidential Agreement

PI professionals' beliefs about likely intervention success change over time and they are responsive to new evidence (information) that is received. RQ1 addressed this question in detail by examining three repeated measures of likely intervention success: Pr_1, Pr_2, and Pr_3. Existing beliefs about general likely intervention success are mostly very positive (M=68.09, SD=16.68). As case-specific information was introduced, second observations of likely success were practically similar (M=69.90) and less variability was demonstrated in these assessments (SD=13.38). This may be related to the instrument development techniques related to the face validity and coherence of scenario item stems, including selection of scenarios based on a high rating of representativeness from field experts and the application of a consistent template format for introducing

each scenario. A notable exception where Pr_2 differed from Pr_1 was Scientific-Supportive Scenario 1 (SS1), which described a performance issue in commissioning a new class of cutter ships for the Navy (see Figure 12).

> **SS1 Scenario -** A national Navy is commissioning a new class of cutter ships. The ship's design and equipment will be somewhat different from the existing cutter ships, but the communications, surveillance, and combat systems will employ brand new technologies. Therefore, none of the ship's crew will have any familiarity or expertise in using it. The ships will be ready in nine months, but before the crew and take possession of the ship from the ship builder they must be capable of fully operating and maintaining it (and the technology it employs).

Figure 12. SS1 scenario.

This scenario elicited more instances of selecting a Knowledge intervention as an initial intervention choice than any other scenario in the study: approximately 90% (*n*=55) of subjects indicated that *well-designed instruction* was most likely to resolve the performance discrepancy. Given the field's well documented basis in education (Brethower, 2008; Rummler, 2007) and instructional systems design (Guerra, 2001b; O'Driscoll, 2003; Reiser, 2001; Rosenberg, Coscarelli, & Hutchinson, 1999; D. T. Tosti & Kaufman, 2007), it follows that PI professionals would be well-versed in identifying situational indicators that a performance problem is suited to Knowledge interventions over other types of interventions. Of note, the same number of subjects that selected *well-designed instruction* for SS1 also marked the highest levels of familiarity for Knowledge interventions: *highly* or *very familiar* (*n*=55, 90.16%).

More generally, pronounced differences occurred between the second and final posterior assessments of likely intervention success, likely due to significant interaction

between time and evidential nature of agreement. Over time, probabilities were approximately 21.48 points lower when infirming evidence was provided. Looking at Pr_3 in particular, group means for scientific and craft-based supportive evidence were 80.33 and 80.83, respectively. Their infirming counterparts were considerably lower (37.95 and 40.83). In fact, a main effect was also noted for evidential agreement. Similarly, the results of the three Factorial MANOVAs performed in RQ2 showed significant effects for evidential agreement and the dependent variables (Pr_3 and the log measure of changes in beliefs between Pr_2 and Pr_3).

In the field of performance improvement, professionals are urged to avoid premature commitment to a particular intervention (Guerra, 2001b; Guerra, 2003; Hybert, 2001; Thomas, 2006, 2007). RQ3 showed that despite revising their beliefs about likely intervention success, professionals tended to stick with their initial intervention choice: observed proportions of sticking were 64.84% in the final study and even higher in the pilot (69.54%).

So far, results are very much in keeping with research on confirmation bias, where supportive data are treated differently than counterevidence (Chapman, 1973) and escalation of commitment, where commitment to a course of action is increased despite negative consequences (Staw, 1976). Despite having been presented with equal proportions of supportive and infirming evidence, subjects tended to stick with their initial choice. These results are probably illustrative of the factors associated with escalation of commitment through meta-analysis (Sleesman et al., 2012). Although

making a private decision, the use of the word *still* in the question may have triggered personal responsibility and ego threat, as modeled in Figure 13.

Do you still think ${q:// SelectedChoices} is most likely to resolve the performance gap with the [performance problem] at [organization name]?
- ❍ Yes (1)
- ❍ No (2)

Figure 13. Stick-or-switch cue.

Furthermore, three subjects responded to an open-ended query at the end of the questionnaire with comments indicating discomfort in identifying a possible solution because of sparse information in the scenarios; thus, uncertainty and having expressed preference for their initial choice may also have contributed to escalated commitment to initial intervention choices.

There was one notable exception where rates of sticking and switching were similar: Craft-Infirming Scenario 5 (AD5), depicted in Figure 14.

AD5 Scenario - Primaria Insurance has offices in 13 states and approximately 650 employees. At Primaria, the marketing and underwriting functions are combined. But, the 90 marketing underwriters take inconsistent approaches to selling and overall premiums are down 25% in the past three years.

AD5 Evidence - It's your first day on site at Primaria and you are meeting with your main client, the Marketing and Underwriting VP and the rest of his management team. They present you with more information about the premium trends over the past years and discuss individual differences among the marketing underwriters. About halfway through the meeting, you mention your initial thoughts about ${q://QID48/ChoiceGroup/SelectedChoices}, when the Marketing and Underwriting VP interrupts, "No, no, no. That's not what we're looking for at all. I want a different kind of solution altogether."

Figure 14. AD5 scenario and evidence.

As shown above, AD5 involved professionals receiving feedback from a key client stakeholder indicating that another type of solution was desired. While prevalent

guidelines for intervention selection tend to emphasize cause-based approaches (Bullock, 1973; Gilbert, 1978; Mager & Pipe, 1997; Darabi, 2003), similar rates of sticking and switching in this case may suggest that PI professionals are evaluating evidence for other probative features (beyond diagnosticity), including the intervention's acceptability to the organization and perceived level of support from its leaders.

Scientific Nature of Evidence

To this point, Chapter 5 has remained silent on the role that the scientific nature of evidence plays in intervention selection decisions. In fact, significant results associated with scientific nature of evidence were rare.

RQ2 looked for differences in persuasion (as measured by Pr_3 and l). A separate MANOVA was conducted for each block of scenarios. Scientific nature of evidence was only significant in Block B and post hoc univariate tests showed only a significant difference for l. Therefore, the significant differences occurred for the amount of change that occurred between Pr_2 and Pr_3 (and not for the posterior probability level).

Since similar effects were not demonstrated for Scientific nature of evidence in RQ1 or Blocks A and B, this result may best be interpreted in light of the scenarios in Block B, which included SS1, Scientific-Infirming Scenario 13 (SD13), Craft-Infirming 2 (AD2), and Craft-Supportive (AS14). Prior and posterior probabilities for all scenarios, including Block B, are illustrated in Figure 15.

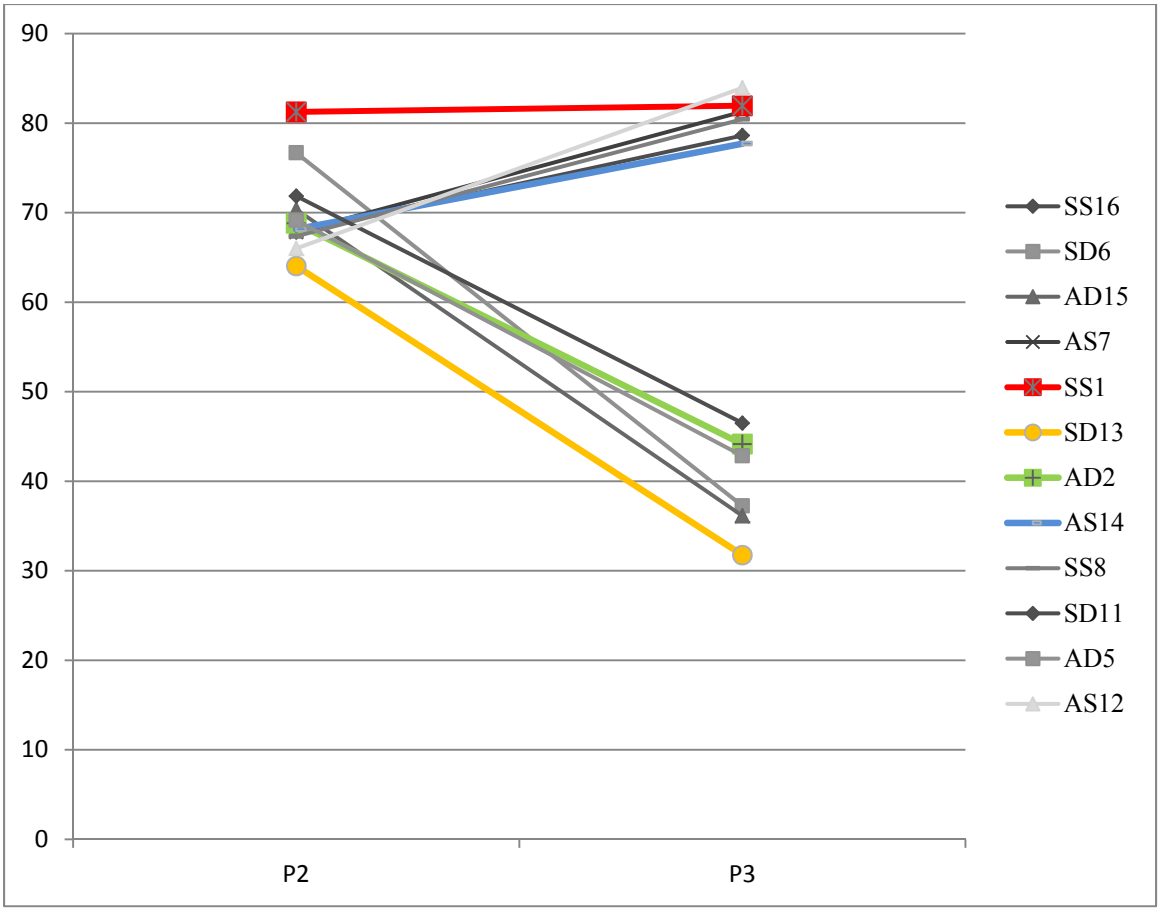


Figure 15. Prior and posterior probabilities for all scenarios.

Note: Scenarios in block B are indicated in color.

Almost no change occurred between Pr_2 and Pr_3 on SS1 and, as previously mentioned, the Pr_2 values for that scenario were high. Meanwhile scenarios AD2 and AS14 exhibited typical interactions and changes. Essentially the negligible difference in SS1, the balanced differences between AD2 and AS14 may have been washed out by the dramatic changes in Pr_2 and Pr_3 for Scientific-Infirming Scenario 13 (SD13). The lack of scientific effect for Block B on Pr_3 in the univariate analysis in RQ2 supports this conclusion.

As shown in Figure 16, the evidence provided in SD13 gave strong counterevidence for the subjects' initial choice, even suggesting that the performance problem is worse after the pilot intervention:

> **SD13 Scenario** - State University's Housing Department interviews, selects, and trains all housing staff, including House Directors and Resident Assistants. This year, there are approximately 50 candidates for open positions and each candidate participates in two one-hour interviews. As a part of Housing Department policy, each current staff member must attend at least one of each candidate's interviews. However, the Housing Director, Andrea Reynolds is frustrated because the interviews are poorly attended by current staff. Moreover, tension is growing in the department because those who do attend the interviews resent that others do not attend even though it is required. Having received a number of complaints, Ms. Reynolds is unsure about what to do to ensure that quality applicants are selected and resolve the conflict among the staff.
>
> **SD13 Evidence** - In your discussions with Ms. Reynolds you share your thoughts about using ${q://QID104/ChoiceGroup/SelectedChoices}. Fortunately, there are several cycles of interviews during the year and your design and development phases align with the upcoming round of interviews. Given this, you and the Housing Director agree to pilot test a prototype of ${q://QID104/ChoiceGroup/SelectedChoices} at that time. You and Ms. Reynolds monitor both attendance and complaints from the staff during the pilot test; unfortunately average attendance is 17% lower at the interviews. The number of complaints increased by 5%, but Ms. Reynolds also has had reports of a major argument about attendance at interviews among some of the tenured staff.

Figure 16. SD 13 scenario and evidence.

Therefore, the significant finding for scientific nature of evidence in Block B may be more of a function of the weightiness of the evidence provided in SD13 (as compared to the other scenarios in the block) than a discourse on differences between scientific and craft-based sources of evidence.

There was one other area where a significant effect for scientific nature of evidence was noted. The frequentist results illustrated a moderate correlation with

sticking to an initial intervention choice (r_s=-0.3160, p<.0001). Bayesian tests noted a significant but almost negligible correlation with switching (r_s=.0283, α=.05, *HPD* Interval =-0.0832, -0.0572). Differences in these results may well be due to the nature of evidence that was provided in the scenarios that were eliminated in the final study: dreaming, personal insight, staff interviews, and the findings of multiple research studies that failed to reject null hypothesis (SD3). Scenario SD3 was removed from the final instrument because *l* measures were inconsistent with other Scientific-Infirming scenarios. In retrospect, the incongruous responsiveness measure on SD3 may have serious implications for conclusions about how scientific evidence is interpreted by PI professionals, especially as it included "fail to reject the hypothesis" terminology.

Familiarity with Interventions

PI professionals should be familiar with a variety of interventions (Hutchison et al., 1996; Hutchison & Stein, 1998; Medsker et al., 1995; Wellins & Rothwell, 2008; Sincoff & Owen, 2004; Watkins, Leigh, and Kaufman, 2000). High levels of familiarity were reported in this study, with the highest levels of familiarity being associated with Knowledge interventions. This may relate to the field's origins in education (Brethower, 2008; Rummler, 2007) and instructional systems design (Guerra, 2001b; O'Driscoll, 2003; Reiser, 2001; Rosenberg, Coscarelli, & Hutchinson, 1999; Tosti & Kaufman, 2007). Further, the reported familiarity with well-designed instruction is not surprising given that Training interventions are given such a high emphasis in performance improvement curricula (Medsker et al., 1995).

Van Tiem (2004) found a significant strong positive correlation between years of experience in the field and expertise with interventions: r=.703, p=.000. Findings in the exploratory analysis for the present study are dissimilar, as only a negligible relationship was observed between years of experience in the field and self-reported familiarity with interventions (r_s=0.0563). Admittedly, this finding is counterintuitive.

A variety of explanations are possible. Van Tiem (2004) measured expertise on a positive to negative semantic differential response scale, including the choices of *expert*, *implements with support, somewhat familiar, aware but little knowledge,* and *do not know*. The present study asked about familiarity on a semantic differential response scale that increased as it was viewed from left to right: *not at all, somewhat, moderately, very,* to *highly familiar*. Additionally, Van Tiem (2004) positioned questions about expertise after questions about the frequency of use for interventions, which may have primed subjects to indicate higher levels of expertise. In addition, Van Tiem (2004) surveyed members of local ISPI and ASTD chapters in southeastern Michigan in 2001; possibly, her findings may be a characteristic of the geographic region. A very likely explanation is that the present study sampled CPTs and those that responded had a high-level, but restricted, range of experience (M=23.630, SD=10.673). Van Tiem (2004) did not report the range of her sample, but lower levels of experience were noted, namely, 66% with eight or more years, and 45% with 10 or more years. In conclusion, if self-reported familiarity and expertise are a single construct (which they may not be), the characteristic may well develop alongside experience in the field and reach a ceiling once more experience is gained.

Still, the correlation between familiarity and expertise was tested ad hoc and the study had been designed to test for the possible influence of familiarity on intervention selection. The ad hoc tests for RQ5 shed some troubling light on this concern. For both initially and subsequently selected interventions, self-reported familiarity was higher for the selected interventions than non-selected interventions. This suggests that whether aware of it or not, PI professionals tend to prefer interventions with which they are more familiar. Although consideration and selection are different activities the private, repeated nature of the present study suggests that professionals may be unintentionally violating ethical obligations to consider solutions in which they are not an expert (Watkins, Leigh, & Kaufman, 2000).

When professionals *do* change their mind about what intervention is preferred, the familiarity levels are lower than their colleagues who stick with their initial choice. This is likely due to having eliminated a familiar option, "switchers" must then select from the remaining options which have lower levels of familiarity.

Unpacking preference and selection from beliefs about likely success, results continue to be problematically mixed. The present sample showed strong positive correlations at Pr_1 for Knowledge and Motive interventions; more moderate relationships were exhibited at Pr_2 for Incentives and Knowledge interventions. Given the background in the field of instructional design, it is unsurprising to see a positive correlation for familiarity on general assessments of likely success and prior probability estimates. These beliefs may be tied to a professional's longevity and persistence in the field; those who felt otherwise may well have left the field long ago.

Another notable fact about the present sample is that no significant correlations were observed at Pr_3. This suggests that at this point, assessments of likely intervention success were more exclusively based on the scenario-at-hand and the evidence that was received. On the other hand, when data from the pilot study were used as informed priors, the results were dramatically different. Out of 18 comparisons, only three did *not* show significant association: Instrumentation (Pr_1), Capacity (Pr_2), and Instrumentation (Pr_3). As shown in Appendix H, the pilot showed stronger associations between self-reported familiarity and assessments of likely intervention success. In the present data, four correlations were noted (and four more illustrated *p*-values that narrowly missed significance). Essentially, the strength and frequency of association in the pilot data outweighed the present sample. Additionally, the Bayesian analysis of RQ1 showed only main effects for agreement and time. It may be that responding behavior played a factor. Both samples were drawn from the same population, but those that took part in the pilot study responded to initial requests to participate. Those that participated in the final phase of the study had seen these same initial requests, but ignored them. Perhaps agreement to participate in a study, Agreement bias, and Self-reported Familiarity bias are associated with another variable not measured in this study.

Changes of Mind

The third analysis in RQ4 sought to predict changes of mind based on *l* and Pr_3. Both variables were predictive of sticking with an intervention. Chi-Square tests for goodness of fit were significant and approximately 75% of the variance in sticking behavior is explained by Pr_3 and *l*. For each unit of increase in Pr_3, there was 8.5% more

chance that subjects would stick with their initial intervention choice. For each half-unit of increase in *l* increased odds of sticking with intervention by almost two times.

Practice in Probabilistic Reasoning

To control for possible carryover effects that can be introduced in repeated measures design, RQ7 tested for order effects. In both the frequentist and Bayesian analyses, no differences in assessments of likely intervention success were noted between assessments collected at the beginning of the questionnaire versus those that were collected subsequently. While this finding counters concerns for potential validity, it may well be that the complex nature of probability requires more time and practice for the effect to develop. Supporting this, four subjects responded to the optional open-ended question at the end of the questionnaire with comments about the uniqueness of the questionnaire. A more troublesome indicator of subjects' probabilistic inexperience, 9.35% of responses to Pr_2 were below 50. Given that subjects were asked to select the intervention that they thought was *most likely to succeed* and the text anchor for 50 on the sliding response scale was *fifty-fifty*, assessments below 50 should indicate that the selected intervention is perceived as not likely to be effective. Therefore, PI professionals may be somewhat naïve with regard to providing probabilistic assessments of intervention success.

Implications for the Field and Practice

During the late 1990s and early 2000s, many authors gravely insisted on an over-reliance on craft-based approaches by PI practitioners; at a minimum, practitioners were viewed to be unsophisticated about scientific evidence, at worst, they were perceived to be openly distrusting of it (Clark & Estes, 1998; Clark & Estes, 2000; Clark & Estes, 2002; Kaufman & Clark, 1999). While late to the controversy, the present study provides empirical data that informs this discussion. In respect to intervention selection decisions made in this scenario-based questionnaire, PI professionals do not appear to overuse craft-based sources of information over scientific evidence. On the contrary, many of the results of analyses in this study indicate that they are used in about the same manner.

That said, the spirit of the concern raised by Clark and Estes (1998, 2000, 2002) is certainly bolstered: as cognitive bias toward evidence that supports one's initial hypothesis and the sublimation of scientific evidence to justify tenacity to a preliminary intervention choice are far cries from the objectivity and rigor demanded in scientific inquiry. Furthermore, it is unlikely that those advocating for scientific-technical approaches to PI would (or should) be satisfied with equal usage of scientific and craft-based sources of evidence. Rather, practitioners are urged to determine what evidence should be obtained based on whatever claim they are trying to support or disprove, central emphasis being placed on the relevance and probative value of the evidence in relationship to the truth of that claim.

Additionally, intervention selection is more than a simple activity that occurs at the end of performance analysis or the beginning of intervention design and development;

rather it is a multi-temporal *decision* process that merits careful consideration by the practitioner. This is especially demonstrated in the present study, where despite low stakes for having made a preliminary intervention choice, subjects tended to stick with their initial choices.

Given the discontinuity between revisions to beliefs about likely intervention success and higher rates of sticking with a preliminary intervention choice, PI practitioners could possibly benefit from application of a decision theoretic. Practitioners should actively reflect on their probabilistic assessments of likely intervention success and check for congruency between these beliefs, their recommendations, and the evidence that they acquire. Moreover, recognizing tendencies to escalate commitment to an initial intervention choice, practitioners are encouraged to identify and seek out sources of evidence that offer the possibility of disconfirming the suitability of an initial intervention choice. In fact, actively endeavoring to disprove one's intervention hypothesis may well obtain more conclusive data that can bolster buy-in and agreement from decision-makers and stakeholders. At a minimum, practitioners should discuss their beliefs and possible recommendations with other PI professionals, perhaps engaging in dialogue with a colleague playing the role of 'devil's advocate'.

Furthermore, the utility of outcome-oriented approaches to intervention selection such as those described by Svenson (2006), Kaufman et al. (1997), and Stolovitch and Keeps (2008) is underscored. Establishing requirements, identifying several means for resolving a performance discrepancy, and evaluating intervention suitability on *a variety of factors* may assist practitioners to avoid bias through systematic identification of trade-

offs. In fact, practitioners are urged to engage in careful consideration of plausible negative consequences. Ideally this would occur *before* recommending a solution, with continued emphasis during the design and development phases of a project.

Beyond the nature of the field and practitioner-level practices, this study does have a few possible implications. For those who oversee the work of PI practitioners, attention should be paid to creating an environment that is conducive to the development of rigorous problem assessment, rational intervention recommendations, and evaluation. Specific guidelines may be drawn from the tactics that are shown to counter the cognitive biases of base-rate neglect, confirmation bias, and escalation of commitment, e.g. collecting baseline data and information, conducting cost-benefit analyses, and accurately framing performance discrepancies. In many respects, supporting a culture of disconfirmation with PI professionals could be simplified by requiring PI professionals to bring the same measure of skepticism to their own intervention hypotheses as they already afford to those offered by a client at the time of request. This point may be heartening to PI clients: not only are these biases also observed in the general population, but extending existing practices may help to alleviate PI professional biases in intervention selection decisions in short measure.

Limitations

As noted in the Introduction there were a few potential limitations. First, use of certified performance technologists as a sampling frame had a potential for introducing sample selection bias. Therefore, the reader is cautioned against generalizing conclusions to instructional designers and performance improvement *practitioners*. As previously

noted, inferences regarding performance improvement *professionals* may be supported by the present study.

In this study, subjects responded to a scenario rather than being observed *in situ*. Although concern may be raised about generalizability to actual practice, these scenarios were developed using guidelines from similar tests that were predictive of on-the-job performance (McDaniel & Nguyen, 2001; Motowidlo, Dunnette, & Carter, 1990; Motowidlo & Tippins, 1993; Weekley & Jones, 1999).

Within-subjects designs may introduce practice effects (Greenwald, 1976). The results of RQ7 suggest that practice effects were not present. However, given the complexity of probabilistic reasoning, it may be that the study was not long enough to observe the effects of practice.

A final possible limitation is that the relatively low response rate in the study (9.06%) may have introduced response bias. As previously noted, this concern is countered by lack of differences between partial and complete responses on the demographics and $Pr_{1,}$ $Pr_{2,}$ and Pr_3 variables. Furthermore, an additional 52 CPTs participated in the pilot phase and mean estimates from this phase were used as informative prior probabilities in the Bayesian analyses. This approach bolsters the validity of the findings and introduces an element of replication to this study.

These limitations do provide some guidance for future research, detailed in the next section.

Future Research

As it studied only CPTs, replication with other PI professional certifications (e.g., CPLP) is recommended. Of course, individuals who pursue a certification may have different characteristics than the general population of PI practitioners. Therefore, it is suggested that similar studies be conducted with the broader population of PI practitioners. Furthermore, given the high level of experience observed in the present sample, comparisons with more novice practitioners are suggested as well.

Despite general scenario descriptions, subjects tended to escalate commitment to a preliminary intervention choice. It is recommended that further study occur around the phrasing of the stick-or-switch cue. This sort of study could help to determine how sensitive escalated commitment is and may even identify ways to reduce it.

Turning from limitations to study findings, the effects of scientific nature of evidence were very limited, but mixed. The main study showed significant, if moderate, correlation between scientific evidence and sticking with an intervention. Bayesian results did not parallel. Therefore, replication of the initial analyses in RQ4 may be worthwhile. Alongside this, findings mostly indicated no effects for scientific nature of evidence, but a significant effect was found in Block B on Pr_3 and *l*. The discussion section addresses possible explanations, including the weightiness of the scientific-infirming scenario in that block and the nature of other scenarios included. Conducting a similar study with re-randomized blocks may well establish more conclusively that Scientific nature of evidence is not a key factor in PI professionals' assessments of likely interventions success. Finally, differences were noted between frequentist and Bayesian

analyses related to Scientific nature of evidence and sticking to an initial intervention. This may have been due to differences in the pilot sample's characteristics or may have been a function of changes between the pilot and final survey instruments. The former included scenarios where the following types of evidence were provided: dreaming, personal insight, staff interviews, and the findings of multiple research studies that failed to reject a null hypothesis. Given the differences between analyses, PI professionals' reactions to these sources of evidence should be studied further. Interpretations of supportive research should be given special attention, as subjects responded inconsistently to evidence where multiple research studies failed to reject null hypothesis.

Also, although no practice effects were exhibited; exploratory analyses and anecdotal reports indicated some naiveté with regard to probabilistic assessments. Longer, possibly longitudinal studies may provide contrary findings to the present study. More generally related to naiveté, it would be practically useful to see if the agreement and escalation to commitment effects noted in the present study would be present with less naïve PI professionals who were educated on subjective probability and cognitive biases. Results of such a study could provide straightforward means for making less biased intervention selection decisions.

The present study introduced a unique perspective on intervention selection, urging a transition from task-oriented guidelines to a multi-temporal decision theoretic. For the first time, intervention selection was linked with probabilistic assessments of intervention success, nature of evidence, and self-reported familiarity. Still, the causal relationships are hardly unpacked and further research involving structural equation

modeling may be of utility in doing so. Modelling of existing and newly-identified variables could provide options for countering some of the biased effects noted here. For instance, are there other variables (besides Pr_3, *l*, and Scientific nature of evidence) that are associated with sticking to an initial intervention choice? What variables (if any) moderate the relationship of self-reported interventions and selected interventions? Additionally, based on the equal rates of switching and sticking for the scenario that involved client feedback, do changes in posterior probability vary dependent on the supportive nature of the client's perspective?

This study involved responses to scenarios. While this method provided the necessary constraints to study the variables involved in the main research questions, it also introduced potential limitations of self-reporting and generalizability to actual practice. As such, more naturalistic studies of intervention selection decisions may be warranted. An additional merit of naturalistic study of intervention selection is that studies of this sort would allow for questions about possible confirmation bias in PI professionals' information search processes, which was not considered here.

Finally, the use of Bayesian statistical methods have been growing exponentially since the 1970s (Berger, 2000), but their use is only just emerging with PI professionals (Pershing et al., 2008b). This study successfully applied Bayes' theorem in its design and statistical analyses. However, further PI study in the Bayesian tradition is merited, as it may help to familiarize practitioners with techniques already used in a variety of relevant areas and elucidate their suitability for analyzing performance improvement problems. An immediate avenue to explore is the few incongruities that were noted between

frequentist and Bayesian analyses in the present study. Therefore, replicating RQ1, the first two analyses in RQ4, and RQ6 is recommended. Finally, this study did not include externally verifiable, frequentist probabilities and the findings are not germane to the normative assertion of Bayes' theorem. Therefore, it is recommended that more controlled experiments be designed to measure the extent to which PI professionals' revisions of assessments of likely intervention success vary from what is called for by Bayes' rule.

Conclusions

Intervention selection is a critical part of the performance improvement process. However, it has often been represented as a task or activity, rather than a complex decision that occurs across time. Alongside this, there has been little research into what sources of evidence are used in intervention selection and what changes in belief occur while performance improvement professionals are making up their mind about which intervention to recommend. Framed from a decision theoretic, this study was an initial step toward resolving this problem. Results from this study showed problematic findings. Probabilistic assessments of likely intervention success were biased by the receipt of evidence that agreed with initial intervention choices. By extension, commitment to these preliminary choices was maintained and even escalated in the face of counterevidence. In the rare cases where PI professionals did change their minds about which intervention they preferred, there were differences in the level of self-reported familiarity with those interventions. This is likely a function of selected interventions having higher levels of self-reported familiarity than their unselected counterparts do. Additionally, while only

limited correlations between self-reported familiarity and assessments of likely intervention success were exhibited in the present sample, inclusion of prior data yielded dramatically different results. Fifteen out of the 18 comparisons performed showed an association between self-reported familiarity and likely intervention success. Despite possible concerns about carryover effects, no practice effects were noted in the study.

These findings bear on the long-standing concern about the technical nature of performance improvement and practitioners are strongly encouraged to approach intervention selection as a decision, rather than a task, with continual attention to congruence in their preferences, beliefs, and the evidence they obtain. Future research with other types of performance improvement practitioners, replication studies, longitudinal, structural equation modeling, externally verifiable probabilities, and natural environments are recommended.

APPENDIX A – INTERVENTION TYPOLOGIES SUMMARY

Typology	No. of intervention classes	Description
Gilbert (1978)	6	Interventions are ways of leveraging change within the stimulus-response-stimulus process at both the individual and environmental levels, resulting in a 2x3 matrix.
Hill & Brethower (1997)	2	Suggests the classification of interventions as either instructional or performance systems (rather than instructional or non-instructional).
Harless as cited in Langdon (1997b)	5	At a simplified level, interventions fall into general classes which may be further sub-divided to aid in selection and management.
Hutchison, Stein & Carleton (1996) and Hutchison and Stein (1998)	20	Defines classes of interventions and particular tactics included in them as a means for identifying the field and a framework for developing expertise as a performance improvement practitioner.
Van Tiem et al.(2001)	7	Refines previous intervention classification systems to reflect current practice in the field of human resource development.
Watkins (2007b; 2003); Watkins and Wedman (2003)	7	Relates classes of interventions to one another for guidance in selecting and aligning multiple interventions.

APPENDIX B – TEXT-BASED VERSION OF QUESTIONNAIRE

Title: Intervention Selection by Performance Improvement Professionals
Principal Investigator: Hillary Leigh

Purpose: As a Certified Performance Technologist, you are being asked to participate in a pilot research study of a questionnaire related to intervention selection. The data from this pilot study will be used to assess the instrument's reliability and to refine the final questionnaire. The research study is being conducted at Wayne State University.

Study Procedures: If you take part in the study, you will be asked to complete a short questionnaire. During this questionnaire, you will be asked about some background demographics and experience. Then you will be asked to assess 12 performance improvement scenarios and the likely success of particular interventions. The questionnaire takes approximately 20 - 25 minutes to complete.

Benefits: As a participant in this research study, there may be no direct benefit for you; however, information from this study may benefit other people now or in the future.

Risks: There are no known risks at this time to participation in this study.

Costs: There will be no costs to you for participation in this research study.

Compensation: You will not be paid for taking part in this study. However, an incentive is being offered to offset the time and inconvenience of participation. 50 participants will be randomly selected for the participant to receive a $25 gift certificate to www.restaurant.com. This gift certificate is redeemable at over 13,000 restaurants in the United States.

Confidentiality: All information collected about you during the course of this study will be kept without any identifiers.

Voluntary Participation /Withdrawal: Taking part in this study is voluntary. You are free to not answer any questions or withdraw at any time. Your decision will not change any present or future relationships with Wayne State University or its affiliates.

Questions: If you have any questions about this study now or in the future, you may contact Hillary Leigh or one of her research team members at the following phone number (562) 331-2332. If you have questions or concerns about your rights as a research participant, the Chair of the Human Investigation Committee can be contacted at (313) 577-1628. If you are unable to contact the research staff, or if you want to talk to someone other than the research staff, you may also call (313) 577-1628 to ask questions or voice concerns or complaints.

Participation: By completing the questionnaire you are agreeing to participate in this study

❍ Proceed to questionnaire (1)

H1 BACKGROUND & DEMOGRAPHICS

Q1 Please indicate your gender
- ❍ Male (1)
- ❍ Female (2)

Q3 Enter your age (in years)

Q4 What is the highest level of education that you have completed?
- ❍ High school or equivalent (1)
- ❍ Associate's degree or certificate (2)
- ❍ Bachelor's degree (3)
- ❍ Master's degree (4)
- ❍ Doctoral degree (5)
- ❍ Other (please list) (6) _____

Q5 Of the following, which industry best represents your work organization:
- ❍ Accommodation and Food Services (1)
- ❍ Administrative and Support and Waste Management and Remediation Services (2)
- ❍ Arts, Entertainment, and Recreation (3)
- ❍ Agriculture, Forestry, Fishing and Hunting (4)
- ❍ Construction (5)
- ❍ Educational Services (6)
- ❍ Health Care and Social Assistance (7)
- ❍ Information (8)
- ❍ Finance and Insurance (9)
- ❍ Management of Companies and Enterprises (10)
- ❍ Manufacturing (11)
- ❍ Mining, Quarrying, and Oil and Gas Extraction (12)
- ❍ Professional, Scientific, Technical Services, and Consulting (13)
- ❍ Public Administration (14)
- ❍ Real Estate and Rental and Leasing (15)
- ❍ Retail Trade (16)
- ❍ Transportation and Warehousing (17)
- ❍ Utilities (18)
- ❍ Wholesale Trade (19)
- ❍ Other (20) _____

Q6 Of the following, which best represents how you would describe your work role:
- ❍ Practitioner (1)
- ❍ Researcher (2)
- ❍ Both (3)

Q7 Indicate the professional societies or associations to which you belong:
- ❍ American Educational Research Association (AERA) (1)
- ❍ American Evaluation Association (AEA) (2)
- ❍ American Society for Training & Development (ASTD) (3)
- ❍ Association for Behavior Analysis International (ABAI) (4)
- ❍ Association for Educational Communications and Technology (AECT) (5)
- ❍ International Society for Performance Improvement (ISPI) (6)
- ❍ Society for Human Resource Management (SHRM) (7)
- ❍ Society for Industrial & Organizational Psychology (SIOP) (8)
- ❍ Other (please list) (9) _____

H2 EXPERIENCE

Q8 How many full years of experience do you have in the field of performance improvement?

Q11 For each type of intervention, please indicate your level of familiarity with it.

	Not at all familiar (1)	Somewhat familiar (2)	Moderately familiar (3)	Very familiar (4)	Highly familiar (5)
Performance data, appraisal, information, policies, and feedback systems (1)	❍	❍	❍	❍	❍
Organizational design, process improvement, resource management, and ergonomics (2)	❍	❍	❍	❍	❍
Financial and non-financial incentives, and career development (3)	❍	❍	❍	❍	❍
Well-designed instruction (4)	❍	❍	❍	❍	❍
Staff scheduling and selection systems (5)	❍	❍	❍	❍	❍
Motivation systems; recruitment of staff that hold organizational values (6)	❍	❍	❍	❍	❍

Q12 Now, thinking about situations in general, how likely do you think it is that each type of intervention is successful in closing a performance gap? Make your selection by adjusting the slider along the bar, where 0=(Almost) impossible, 50=fifty-fifty, and 100= (Almost certain). As you drag the slider, the value you have selected will appear in bold to the right.

_____ Performance data, appraisal, information, policies, and feedback systems (1)

_____ Organizational design, process improvement, resource management, and ergonomics (2)

_____ Financial and non-financial incentives, and career development (3)

_____ Well-designed instruction (4)

_____ Staff scheduling and selection systems (5)

_____ Motivation systems; recruitment of staff that hold organizational values (6)

H3 Now you will respond to a few scenarios. In each case you will get some background information about a gap in performance. Based on that introduction, you will indicate which intervention you prefer and your assessment of how likely it is to resolve the gap in performance. After getting additional information, you will complete these steps again.

SD6S Kayak Adventures is a kayak tour and surf lesson business with a handful of locations along the coast of South Carolina. The company offers eco tours, guided and self-guided tours, as well as camping at remote sites. A tour can be booked in person, by phone, or online. In order to maintain a safe environment for tour guides and customers, overbooking of tours is avoided. However, multiple people manage the booking process and in the last 3 months all of the tours have either been overbooked (by as much as 50%) or severely underbooked (resulting in cancellations). Since this problem impacts both revenue and safety, the company wants to resolve the issue quickly.

SD6II Which intervention is most likely to resolve the performance gap?
- performance data, appraisal, information, policies, and feedback systems (1)
- organizational design, process improvement, resource management, and ergonomics (2)
- financial and non-financial incentives, and career development (3)
- well-designed instruction (4)
- staff scheduling and selection systems (5)
- motivation systems; recruitment of staff that hold organizational values (6)

SD6P2 How likely is ${q://QID55/ChoiceGroup/SelectedChoices} to resolve the performance gap with the booking process at Kayak Adventures?
_____ ${q://QID55/ChoiceGroup/SelectedChoices} (1)

SD6E On your first day working with Kayak Adventures, you meet with all of the staff who manages the booking process. In addition to walking through the steps for booking in person, phone, or by web you learn some of the past history of the process. Through discussions with these subject matter experts (SMES), you learn that Kayak Adventures implemented ${q://QID55/ChoiceGroup/SelectedChoices} programs last year. In fact, the booking process SMEs report that not only was the program implemented, but there was no effect. The problems with overbooking and under booking continued at the same rates as before.

SD6P3 Based on this, now how likely is ${q://QID55/ChoiceGroup/SelectedChoices} to resolve the performance gap with the booking process at Kayak Adventures?
_____ ${q://QID55/ChoiceGroup/SelectedChoices} (1)

SD6X Do you still think ${q://QID55/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap with the booking process at Kayak Adventures?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you still think ${q://QID50/ChoiceGroup/Selected... No Is Selected

SD62I If not, which intervention is most likely to resolve the performance gap with Primaria's Marketing Underwriters?

AD15S Lumis Group manages a chain of budget, extended stay, and luxury hotels. On an annual basis, it conducts an employee satisfaction survey. Last year's results indicate a problem with the performance evaluation process, as only a small percentage (34%) indicated that performance evaluations were fair. Concerned about the widespread nature of this perception and possible legal implications, Human Resources would like to design a new performance evaluation program.

AD15II Which intervention is most likely to resolve the performance gap at Lumis?
- ❍ performance data, appraisal, information, policies, and feedback systems (1)
- ❍ organizational design, process improvement, resource management, and ergonomics (2)
- ❍ financial and non-financial incentives, and career development (3)
- ❍ well-designed instruction (4)
- ❍ staff scheduling and selection systems (5)
- ❍ motivation systems; recruitment of staff that hold organizational values (6)

AD15P2 How likely is ${q://QID119/ChoiceGroup/SelectedChoices}to resolve the performance gap at Lumis?
_____ ${q://QID119/ChoiceGroup/SelectedChoices} (1)

AD15E You return to your office to summarize your onsite observations. As you are finishing up a colleague, Dave, drops in and asks about the work with Lumis. You take him through your notes, including your initial thoughts about using ${q://QID119/ChoiceGroup/SelectedChoices}, documenting it on a flip chart. A long-time collaborator, Dave rolls up his sleeves and takes the marker from you. He draws a circle around one of the findings from your observations, and asks, "But what about this?" You shrug and concede, "Yeah, it doesn't address that point at all…" He nods and draws another circle, and then another; the two of you brainstorm for a while more. As you step back to assess the marked up flip chart, you both remark, "So it looks like ${q://QID119/ChoiceGroup/SelectedChoices} isn't the way to go."

AD15P3 Based on this, now how likely is ${q://QID119/ChoiceGroup/SelectedChoices} to resolve the performance gap at Lumis?
_____ ${q://QID119/ChoiceGroup/SelectedChoices} (1)

AD15X Do you still think ${q://QID119/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap at Lumis?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you still think ${q://QID119/ChoiceGroup/Se... No Is Selected

AD152I If not, which intervention is most likely to resolve the performance gap at Lumis?

AS7S Frontier Healthcare Association (FHA) is 501(c)(3) nonprofit organization that includes a network of 56 community health clinics in rural communities across Montana, Wyoming, the Dakotas, and Nebraska. Recently, this network of clinics began expanding and has seen double digit annual increases in membership. As a result, there is an increased demand for Information Technology (IT) and other technical support. You were engaged by FHA's Board of Directors, with the goal of increasing satisfaction, usage, and suitability of their technical support.

AS7II Which intervention is most likely to resolve the performance gap?
- ❍ performance data, appraisal, information, policies, and feedback systems (1)
- ❍ organizational design, process improvement, resource management, and ergonomics (2)
- ❍ financial and non-financial incentives, and career development (3)
- ❍ well-designed instruction (4)
- ❍ staff scheduling and selection systems (5)
- ❍ motivation systems; recruitment of staff that hold organizational values (6)

AS7P2 How likely is ${q://QID62/ChoiceGroup/SelectedChoices} to resolve the performance gap at FHA?
_____ ${q://QID62/ChoiceGroup/SelectedChoices} (1)

AS7E To help you prepare to meet with FHA's Board, you do some internet research on community health clinics, rural health care delivery, and technical support. You review a number of websites, wikis, and blogs that outline ${q://QID62/ChoiceGroup/SelectedChoices} programs. The sites describe increases in satisfaction and utilization of technical support when these ${q://QID62/ChoiceGroup/SelectedChoices} programs are applied.

AS7P3 Based on this, now how likely is ${q://QID62/ChoiceGroup/SelectedChoices} to resolve the performance gap at FHA?
_____ ${q://QID62/ChoiceGroup/SelectedChoices} (1)

AS7X Do you still think ${q://QID62/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap at FHA?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you still think ${q://QID62/ChoiceGroup/Sel... No Is Selected

AS72I If not, which intervention is most likely to resolve the performance gap at FHA?

SS16S Mt. Jones General is a 240 bed community hospital in a major metropolitan area. With 45 beds, the Cardiac Telemetry unit provides monitoring and care for patients who have had or at risk for cardiac events. Recently, the unit has been experiencing a problem with patient falls. With any fall, there is a risk for injury, and the unit manager is concerned about both increased costs and patient safety.

SS16II Which intervention is most likely to resolve the performance gap at Mt. Jones General's Cardiac Telemetry unit?
- ❍ performance data, appraisal, information, policies, and feedback systems (1)
- ❍ organizational design, process improvement, resource management, and ergonomics (2)
- ❍ financial and non-financial incentives, and career development (3)
- ❍ well-designed instruction (4)
- ❍ staff scheduling and selection systems (5)
- ❍ motivation systems; recruitment of staff that hold organizational values (6)

SS16P2 How likely is ${q://QID126/ChoiceGroup/SelectedChoices} to resolve the performance gap at Mt. Jones General's Cardiac Telemetry unit?
_____ ${q://QID126/ChoiceGroup/SelectedChoices} (1)

SS16E You schedule a session to do a cause analysis with major stakeholders and performers. During the session, you work with the unit manager, physicians, and nurses to construct a fishbone diagram showing a graphical representation of causes for falls on the unit, that reveal causes in people, methods, machines, materials, measurements, and environment. The output of this process noted lack of ${q://QID126/ChoiceGroup/SelectedChoices} as a contributing cause.

SS16P3 Based on this, now how likely is ${q://QID126/ChoiceGroup/SelectedChoices} to resolve the performance gap at Mt. Jones General's Cardiac Telemetry unit?
_____ ${q://QID126/ChoiceGroup/SelectedChoices} (1)

SS16X Do you still think ${q://QID126/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap at Mt. Jones General?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you still think ${q://QID126/ChoiceGroup/Se... No Is Selected

SS162I If not, which intervention is most likely to resolve the performance gap at Mt. Jones General?

SS1S A national Navy is commissioning a new class of cutter ships. The ship's design and equipment will be somewhat different from the existing cutter ships, but the communications, surveillance, and combat systems will employ brand new technologies. Therefore, none of the ship's crew will have any familiarity or expertise in using it. The ships will be ready in nine months, but before the crew and take possession of the ship from the ship builder they must be capable of fully operating and maintaining it (and the technology it employs).

SS1II Which intervention is most likely to resolve the performance gap?
- performance data, appraisal, information, policies, and feedback systems (1)
- organizational design, process improvement, resource management, and ergonomics (2)
- financial and non-financial incentives, and career development (3)
- well-designed instruction (4)
- staff scheduling and selection systems (5)
- motivation systems; recruitment of staff that hold organizational values (6)

SS1P2 How likely is ${q://QID20/ChoiceGroup/SelectedChoices} to resolve the performance gap with your Navy client?
_____ ${q://QID20/ChoiceGroup/SelectedChoices} (1)

SS1E Preparing for your first meeting with your Navy client, you review the theoretical support for ${q://QID20/ChoiceGroup/SelectedChoices}. There is a long history of empirical research that illustrates a strong, positive relationship between ${q://QID20/ChoiceGroup/SelectedChoices} and the development of new skills. Multiple regression models include a variety of factors (e.g. complexity of systems, learner experience, aptitude), yet 68% of the variance in the time in which expertise is established is explained by use of ${q://QID20/ChoiceGroup/SelectedChoices}.

SS1P3 Based on this, now how likely is ${q://QID20/ChoiceGroup/SelectedChoices} to resolve the performance gap with your Navy client?
_____ ${q://QID20/ChoiceGroup/SelectedChoices} (1)

SS1X Do you still think ${q://QID20/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap with your Navy client?
- Yes (1)
- No (2)

Answer If Do you still think ${q://QID20/ChoiceGroup/SelectedC... No Is Selected

SS12I If not, which intervention is most likely to resolve the performance gap with your Navy client?

SD13S State University's Housing Department interviews, selects, and trains all housing staff, including House Directors and Resident Assistants. This year, there are approximately 50 candidates for open positions and each candidate participates in two one-hour interviews. As a part of Housing Department policy, each current staff member must attend at least one of each candidate's interviews. However, the Housing Director, Andrea Reynolds is frustrated because the interviews are poorly attended by current staff. Moreover, tension is growing in the department because those who do attend the interviews resent that others do not attend even though it is required. Having received a number of complaints, Ms. Reynolds is unsure about what to do to ensure that quality applicants are selected and resolve the conflict among the staff.

SD13II Which intervention is most likely to resolve the performance gap in the Housing Department?
- ❍ performance data, appraisal, information, policies, and feedback systems (1)
- ❍ organizational design, process improvement, resource management, and ergonomics (2)
- ❍ financial and non-financial incentives, and career development (3)
- ❍ well-designed instruction (4)
- ❍ staff scheduling and selection systems (5)
- ❍ motivation systems; recruitment of staff that hold organizational values (6)

SD13P2 How likely is ${q://QID104/ChoiceGroup/SelectedChoices} to resolve the performance gap in the Housing Department?
_____ ${q://QID104/ChoiceGroup/SelectedChoices} (1)

SD13E In your discussions with Ms. Reynolds you share your thoughts about using ${q://QID104/ChoiceGroup/SelectedChoices}. Fortunately, there are several cycles of interviews during the year and your design and development phases align with the upcoming round of interviews. Given this, you and the Housing Director agree to pilot test a prototype of ${q://QID104/ChoiceGroup/SelectedChoices} at that time. You and Ms. Reynolds monitor both attendance and complaints from the staff during the pilot test; unfortunately average attendance is 17% lower at the interviews. The number of complaints increased by 5%, but Ms. Reynolds also has had reports of a major argument about attendance at interviews among some of the tenured staff.

SD13P3 Based on this, now how likely is ${q://QID104/ChoiceGroup/SelectedChoices} to resolve the performance gap in the Housing Department?
_____ ${q://QID104/ChoiceGroup/SelectedChoices} (1)

SD13X Do you still think ${q://QID104/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap in the Housing Department?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you still think ${q://QID104/ChoiceGroup/SelectedChoic... No Is Selected

SD132I If not, which intervention is most likely to resolve the performance gap in the Housing Department?

AD2S DLH is a global technology firm that produces and delivers services in almost 150 countries. Its call center provides front-line customer support and manages approximately 2.5 million calls per month. However, when compared with other technology firms of its size, customer satisfaction with the call center is quite low. Customers are on hold for long periods of time, are transferred to multiple departments, and have difficulty learning about new products and services that are available.

AD2II Which intervention is most likely to resolve the performance gap?
- performance data, appraisal, information, policies, and feedback systems (1)
- organizational design, process improvement, resource management, and ergonomics (2)
- financial and non-financial incentives, and career development (3)
- well-designed instruction (4)
- staff scheduling and selection systems (5)
- motivation systems; recruitment of staff that hold organizational values (6)

AD2P2 How likely is ${q://QID26/ChoiceGroup/SelectedChoices} to resolve the performance gap in DLH's call center customer satisfaction?
_____ ${q://QID26/ChoiceGroup/SelectedChoices} (1)

AD2E Once onsite at DLH's call center you meet with the call center manager and floor supervisors. After reviewing customer complaint records, you talk with a few customer service representatives. In the afternoon, you sit in with the floor supervisors and listen to calls as the representatives field them. All day long, you have a nagging feeling…you can't quite put your finger on it…but your intuition tells you your first instincts were wrong and you just have a hunch that ${q://QID26/ChoiceGroup/SelectedChoices} isn't going to work in this situation.

AD2P3 Based on this, now how likely is ${q://QID26/ChoiceGroup/SelectedChoices} to resolve the performance gap in DLH's call center customer satisfaction?
_____ ${q://QID26/ChoiceGroup/SelectedChoices} (1)

AD2X Do you still think ${q://QID26/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap in DLH's call center customer satisfaction?
- Yes (1)
- No (2)

Answer If Do you still think ${q://QID26/ChoiceGroup/SelectedChoice... No Is Selected

AD2I If not, which intervention is most likely to resolve the performance gap in DLH's call center customer satisfaction?

AS14S SDA is a national oversight association for US securities firms. The standards and regulations for securities are highly complex and ever-changing. Additionally, SDA has seen tremendous growth in the last several decades and is now the leading regulatory agency in the country, with responsibility for approximately 5,000 securities examiners. Foreseeing increased demand in the future, SDA would like to train new examiners more quickly (it currently takes novice examiners almost two years to become proficient in their work).

AS14II Which intervention is most likely to resolve the performance gap at SDA?
- performance data, appraisal, information, policies, and feedback systems (1)
- organizational design, process improvement, resource management, and ergonomics (2)
- financial and non-financial incentives, and career development (3)
- well-designed instruction (4)
- staff scheduling and selection systems (5)
- motivation systems; recruitment of staff that hold organizational values (6)

AS14P2 How likely is ${q://QID112/ChoiceGroup/SelectedChoices} to resolve the performance gap at SDA?
_____ ${q://QID112/ChoiceGroup/SelectedChoices} (1)

AS14E As you prepare to meet with SDA, you do some additional research and come across an editorial article by a well-known performance improvement expert describing ${q://QID112/ChoiceGroup/SelectedChoices}. In addition to discussing some historical background, the author presents the general merits of ${q://QID112/ChoiceGroup/SelectedChoices}. The crux of the author's argument is that they are effective in getting novices up to speed quickly: "${q://QID112/ChoiceGroup/SelectedChoices} work especially well in situations where a high-level of expertise is required in order to perform basic job responsibilities, such as those that depend on technical know-how and complex requirements." Surprisingly, much of the story depends on the author's experience with onboarding new employees at another regulatory agency.

AS14P3 Based on this, now how likely is ${q://QID112/ChoiceGroup/SelectedChoices} to resolve the performance gap at SDA?
_____ ${q://QID112/ChoiceGroup/SelectedChoices} (1)

AS14X Do you still think ${q://QID112/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap at SDA?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you still think ${q://QID112/ChoiceGroup/Se... No Is Selected

AS142I If not, which intervention is most likely to resolve the performance gap at SDA?

SD11S Modula is a small pharmaceutical company that employs approximately 45,000 scientific, sales, marketing, administration, regulatory, and manufacturing professionals. There are multiple levels of management in the organization's structure, with up to six or seven levels between front-line staff and executives. Unfortunately, mid-level administrators report that they don’t have a handle on performance management, especially as it relates to employee development and project status/progress.

SD11II Which intervention is most likely to resolve the performance gap?
- ❍ performance data, appraisal, information, policies, and feedback systems (1)
- ❍ organizational design, process improvement, resource management, and ergonomics (2)
- ❍ financial and non-financial incentives, and career development (3)
- ❍ well-designed instruction (4)
- ❍ staff scheduling and selection systems (5)
- ❍ motivation systems; recruitment of staff that hold organizational values (6)

SD11P2 How likely is ${q://QID90/ChoiceGroup/SelectedChoices} to resolve the performance gap at Modula?
_____ ${q://QID90/ChoiceGroup/SelectedChoices} (1)

SD11E As a part of your analysis into the problems with project management at Modula, you request additional information from the mid-level administrators. Each administrator emails you and their direct

reports, who respond to you both work plans for current projects. As you review these email responses, you note that ${q://QID90/ChoiceGroup/SelectedChoices} are alluded to frequently. Further analysis of the content in the work plans shows that ${q://QID90/ChoiceGroup/SelectedChoices} are already in place at Modula; in fact, it is the most frequently mentioned tactic across all of the work plans. Still, the mid-level administrators are not fully aware of project progress.

SD11P3 Based on this, now how likely is ${q://QID90/ChoiceGroup/SelectedChoices} to resolve the performance gap at Modula?
_____ ${q://QID90/ChoiceGroup/SelectedChoices} (1)

SD11X Do you still think ${q://QID90/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap at Modula?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you still think ${q://QID91/ChoiceGroup/Selected... No Is Selected

SD112I If not, which intervention is most likely to resolve the performance gap at Modula?

AS12S Historically, the Continuing Education extension (CEE) program at East Coast State University (ECSU) has maintained a separate student records and registration system. The university's Records and Registration department is switching to a new system and the CEE is being required to implement this same system. When the Records and Registration looked into system options, they projected a cost savings after one year. Still, CEE's senior leaders are very concerned about the time and resources that will be required to adapt the existing processes to the new system.

AS12II Which intervention is most likely to resolve the performance gap at CEE?
- ❍ performance data, appraisal, information, policies, and feedback systems (1)
- ❍ organizational design, process improvement, resource management, and ergonomics (2)
- ❍ financial and non-financial incentives, and career development (3)
- ❍ well-designed instruction (4)
- ❍ staff scheduling and selection systems (5)
- ❍ motivation systems; recruitment of staff that hold organizational values (6)

AS12P2 How likely is ${q://QID97/ChoiceGroup/SelectedChoices} to resolve the performance gap at CEE?
_____ ${q://QID97/ChoiceGroup/SelectedChoices} (1)

AS12E As you reflect on CEE's situation, you recall a comparable system integration project that you worked on for a local community college. The community college faced a similar issue when the system of community colleges it belonged to was transitioning their records and registration system and leaders were concerned about the time and resources to ramp up the system. At the time, you developed and implemented ${q://QID97/ChoiceGroup/SelectedChoices} for them. From your recollection, the evaluation you conducted showed very positive results; an approximate return-on-investment of 300% in two years.

AS12P3 Based on this, now how likely is ${q://QID97/ChoiceGroup/SelectedChoices} to resolve the performance gap at CEE?
_____ ${q://QID97/ChoiceGroup/SelectedChoices} (1)

AS12X Do you still think ${q://QID97/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap at CEE?
- ❍ Yes (1)
- ❍ No (2)

Answer If Do you still think ${q://QID91/ChoiceGroup/Selected... No Is Selected

AS122I If not, which intervention is most likely to resolve the performance gap at CEE?

SS8S One of many accreditation guidelines for health maintenance organizations (HMOs) is that they provide Primary Care Physicians (PCPs) with performance data. This data should measure individual PCPs' performance on a number of utilization, pharmacy, and quality measures and provide the PCP with actionable information to improve their performance in these areas. However, Alliance Health Plan is contacting you because over the past two years reports are being delivered later and later. In fact, the most recent batch exceeded contract requirements by almost 200%.

SS8II Which intervention is most likely to resolve the performance gap?
- ❍ performance data, appraisal, information, policies, and feedback systems (1)
- ❍ organizational design, process improvement, resource management, and ergonomics (2)
- ❍ financial and non-financial incentives, and career development (3)
- ❍ well-designed instruction (4)
- ❍ staff scheduling and selection systems (5)
- ❍ motivation systems; recruitment of staff that hold organizational values (6)

SS8P2 How likely is ${q://QID69/ChoiceGroup/SelectedChoices} to resolve the performance gap with Alliance Health Plan's performance reports?
_____ ${q://QID69/ChoiceGroup/SelectedChoices} (1)

SS8E To get additional information, you search the research literature to see what more you can learn about possible solutions. Coincidentally, you find a case study that was published earlier this year by one of Alliance Health Plan's competitors. As you review the abstract for the case, you see that the competitor was having similar difficulties in meeting the contract requirements for delivering PCP performance reports. As a result, the competitor implemented ${q://QID69/ChoiceGroup/SelectedChoices} programs; a summative evaluation conducted one-year after the implementation showed cycle time decreases ranging from 23% - 68%, with an average of 51% reduced cycle time.

SS8P3 Based on this, now how likely is ${q://QID69/ChoiceGroup/SelectedChoices} to resolve the performance gap with Alliance Health Plan's performance reports?
_____ ${q://QID69/ChoiceGroup/SelectedChoices} (1)

SS8X Do you still think ${q://QID69/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap with Alliance Health Plan's performance report?
- Yes (1)
- No (2)

Answer If Do you still think ${q://QID69/ChoiceGroup/SelectedC... No Is Selected

SS82I If not, which intervention is most likely to resolve the performance gap with Alliance Health Plan's performance report?

AD5S Primaria Insurance has offices in 13 states and approximately 650 employees. At Primaria, the marketing and underwriting functions are combined. But, the 90 marketing underwriters take inconsistent approaches to selling and overall premiums are down 25% in the past three years.

AD5II Which intervention is most likely to resolve the performance gap?
- performance data, appraisal, information, policies, and feedback systems (1)
- organizational design, process improvement, resource management, and ergonomics (2)
- financial and non-financial incentives, and career development (3)
- well-designed instruction (4)
- staff scheduling and selection systems (5)
- motivation systems; recruitment of staff that hold organizational values (6)

AD5P2 How likely is ${q://QID48/ChoiceGroup/SelectedChoices} to resolve the performance gap with Primaria's Marketing Underwriters?
_____ ${q://QID48/ChoiceGroup/SelectedChoices} (1)

AD5E It's your first day on site at Primaria and you are meeting with your main client, the Marketing and Underwriting VP and the rest of his management team. They present you with more information about the premium trends over the past years and discuss individual differences among the marketing underwriters. About halfway through the meeting, you mention your initial thoughts about ${q://QID48/ChoiceGroup/SelectedChoices}, when the Marketing and Underwriting VP interrupts, "No, no, no. That's not what we're looking for at all. I want a different kind of solution altogether."

AD5P3 Based on this, now how likely is ${q://QID48/ChoiceGroup/SelectedChoices} to resolve the performance gap with Primaria's underwriters?
_____ ${q://QID48/ChoiceGroup/SelectedChoices} (1)

AD5X Do you still think ${q://QID48/ChoiceGroup/SelectedChoices} is most likely to resolve the performance gap with Primaria's Marketing Underwriters?
- Yes (1)
- No (2)

Answer If Do you still think ${q://QID48/ChoiceGroup/Sel... No Is Selected

AD52I If not, which intervention is most likely to resolve the performance gap with Primaria's Marketing Underwriters?

Q129 Thank you for completing the questionnaire! Click the button below to submit your responses and proceed to the next page.

Q126 If you choose, please provide any comments related to your assessments or the questionnaire in general:

APPENDIX C – CRITICAL INCIDENT QUESTIONNAIRE

Questionnaire

Background: The purpose of this questionnaire is to gather realistic descriptions of the early phases of performance improvement leading up to intervention selection. Your response will remain completely anonymous.

Please write a true story about your work as a performance improvement professional. Pick a situation that you have encountered or where you have been called in to consult on a performance problem. Here are few questions to guide you:

1. Provide some background about the organization's industry, purpose, employees, etc. and describe the situation leading up to your involvement. What kind of background information did the client provide to you?
2. Why was the performance problem important? How did the problem relate to the goals of the organization, department, or team?
3. If you collected data, what kind of data did you collect (and from what sources)? What more did you learn about the situation? Ultimately, what do you think was "going on" in this situation?
4. What interventions seemed like they might be appropriate? What intervention(s) did you recommend? What intervention(s) were implemented (if any)?

APPENDIX D– EXPERTS' RANK ORDERING OF SCENARIOS

	Percentage of Experts Who Marked Highest Level	
Scenario	Representativeness	Clarity
Lumis Group manages a chain of budget, extended stay, and luxury hotels. On an annual basis, it conducts an employee satisfaction survey. Last year's results indicate a problem with the performance evaluation process, as only a small percentage (34%) indicated that performance evaluations were fair. Concerned about the widespread nature of this perception and possible legal implications, Human Resources would like to design a new performance evaluation program.	82%	60%
Tonacon is an international development organization providing assistance in more than 100 countries with a total of 10,000 human resources employees. Unfortunately, last year's internal audit for the staff training and development (T&D) program was less than favorable. Despite the fact that almost $80M was spent on T&D year, clear linkages to returns on this investment do not exist in most cases. For about a third of the programs, satisfaction data exist. But results beyond that are not captured and the value added by T&D has not been quantified.	82%	60%
Adwell Promotions produces a small classified ad paper. Recently, the Customer Service manager, Amy Lin, has observed an increase in errors related to credit card processing, customer information, and ad verification. Amy is especially concerned because customer complaints are on the rise. Moreover, resolving these errors requires additional time/work and often means lost revenue.	80%	44%
A national Navy is commissioning a new class of cutter ships. Almost every system on these ships will be new, so none of the ship's crew will have any expertise in using it. Before the crew and take possession of the ship from the ship builder, they must be capable of fully operating and maintaining it (and the technology it employs).	73%	50%

Modula is a small pharmaceutical company that employs approximately 45,000 scientific, sales, marketing, administration, regulatory, and manufacturing professionals. Unfortunately, senior administrators report that they don't have a handle on performance management, especially as it relates to employee development and project status/progress.	73%	40%
Mt. Jones General is community hospital in a major metropolitan area. Altogether, the hospital has 240 beds; however one of its units has been experiencing a problem with patient falls. With any fall, there is a risk for injury, and the unit manager is concerned about both increased costs and patient safety.	70%	44%
One of many accreditation guidelines for health maintenance organizations (HMOs) is that they provide Primary Care Physicians (PCPs) with performance data. This data should measure individual PCPs' performance on a number of utilization, pharmacy, and quality measures and provide the PCP with actionable information to improve their performance in these areas. However, Alliance Health Plan is contacting you because over the past two years reports are being delivered later and later. In fact, the most recent batch exceeded contract requirements by almost 200%.	70%	44%
DLH is a global technology firm that produces and delivers services in almost 150 countries. Its call center provides front-line customer support and manages approximately 2.5 million calls per month. However, when compared with other technology firms of its size, customer satisfaction with the call center is quite low. Customers are on hold for long periods of time, are transferred to multiple departments, and have difficulty learning about new products and services that are available.	64%	60%

Kayak Adventures is a kayak tour and surf lesson business with a handful of locations along the coast of South Carolina. The company offers eco tours, guided and self-guided tours, as well as camping at remote sites. A tour can be booked in person, by phone, or online. In order to maintain a safe environment for tour guides and customers, overbooking of tours is avoided. However, multiple people manage the booking process and lately tours have either been overbooked or severely underbooked. Since this problem impacts both revenue and safety, the company wants to resolve the issue quickly.	64%	50%
Historically, the Continuing Education extension (CEE) program at East Coast State University (ECSU) has maintained a separate student records and registration system. The university's Records and Registration department is switching to a new system and the CEE was offered the opportunity to implement this same system. This new system will result in a cost savings, but CEE's senior leaders are concerned about the time and resources that will be required to adapt the existing processes to the new system.	64%	40%
SDA is a national oversight association for US securities firms. The standards and regulations for securities are highly complex and ever-changing. Additionally, SDA has seen tremendous growth in the last several decades and is now the leading regulatory agency in the country, with responsibility for approximately 5,000 securities examiners. Foreseeing increased demand in the future, SDA would like to train new examiners more quickly (it currently takes novice examiners almost two years to become proficient in their work).	64%	40%
Primaria Insurance has offices in 13 states and approximately 650 employees. At Primaria, the marketing and underwriting are functions are combined. But, the 90 Marketing Underwriters take inconsistent approaches to selling and overall premiums are down 25% in the past three years.	60%	44%

Reliable Advice Services provides sound healthcare advice from physicians and nurses. Medical groups contract with Reliable to provide after-hours answering services to their patients. Since many calls that come in after-hours are often urgent in nature, it is expected that calls are returned in 30 minutes or less. However, recent reports indicate that average turnaround time for calls is running closer to 45 minutes.	55%	60%
Frontier Healthcare Association (FHA) is a network of 56 community health clinics in rural communities across Montana, Wyoming, the Dakotas, and Nebraska. Recently, this network of clinics began expanding and as a result of membership growth, there is an increased demand for quality technical assistance services. You were engaged by FHA's Board of Directors, with the goal of increasing satisfaction, usage, and suitability of their technical assistance services.	55%	40%
State University's Housing Department interviews, selects, and trains all housing staff, including House Directors and Resident Assistants. This year, there are approximately 50 candidates for open positions and each candidate participates in two interviews. As a part of Housing Department policy, each current staff member must attend at least one of each candidate's interviews. However, the Housing Director, Andrea Reynolds is frustrated because the interviews are poorly attended by current staff. Moreover, tension is growing in the department because those who do attend the interviews resent that others do not attend even though it is required. Having received a number of complaints, Ms. Reynolds is unsure about what to do to ensure that quality applicants are selected and resolve the conflict among the staff.	55%	40%
WJA Industrial, Inc. is an automotive supplier that supplies parts directly to the major US automakers. Within the Purchasing function of the company, there are approximately 150 employees. Unfortunately, these employees do not fully understand the organizational goals and the Purchasing Director believes that purchasing staff are unaware of both the organizational and executive level objectives	55%	11%

The Development Disabilities Support Network (DDSN) is an agency provides monitoring, support, and funding services to adults with development disabilities. Much of this work is accomplished via contracted services from vendors. Complaints from client and their families about these vendors have risen and there is great inconsistency among these vendors in terms of program practices, staff performance, and capabilities.	50%	33%
Since the early 1990s, Union Memorial Health System has sent a satisfaction survey to patients seen in outpatient clinics. In addition to collecting feedback about appointment scheduling, access, receptionist courtesy, lab services, the survey mainly focuses on satisfaction with healthcare providers' communication. Despite sending monthly reports to the individual providers and their supervisors, performance has remained fairly static. Also, Union Memorial has not improved on the external benchmarks of satisfaction with provider communication and has even fallen behind other health systems in the same region.	50%	33%
Pinnacle Group runs a network of clinics in southern California. Like many healthcare systems in the state, the group struggles with appointment access or "getting an appointment as soon as it is wanted". Unlike many of its competitors, survey responses and complaints from Pinnacle's patients place the group in the bottom quartile for patient satisfaction in this area.	50%	33%
Solarus is a US-based company that manufactures semi-conductors, the materials used in almost all modern electronics. When a new production employee starts working, they are oriented by Production Trainers who teach them about company values, product standards, time management, and product output expectations levels. However, the manufacturing managers have raised concerns with the training department about the production trainers who were recently promoted. Overall there has been a gradual decrease in production output, and there were reports of communication problems, conflict resolution issues, and a lack of unity with the Production Trainers.	50%	22%

Oasis is a small communications firm operating in the Cotton Belt states. Due to exponential growth in the data communications market over the past several years, there is great inconsistency in the Account Executives sales for digital products. Moreover, there is a great disparity between sales in this area compared to Oasis' voice communication products. Recognizing convergence of voice and data communications in an "all-digital" world, Oasis is concerned about the inconsistent and low performance of the Account Executives digital sales performance.	50%	11%
Information Technology Application Development (ITAD) designs, develops, and supports internal IT solutions for a US clothing retailer. A number of projects that ITAD is working on are at least 6 months behind schedule and clients are dissatisfied by the paperwork requirements for IT solutions. Even more problematically, last year's IT project costs exceeded estimates by almost $1.2 million.	45%	60%
You were recently contacted by Marilyn Cowell, who is the Director for the Purchasing department. She is troubled about an ongoing issue she is having with one of the employees that reports to her. This employee is missing deadlines for her tasks, regularly arrives tardy for work and has even failed to attend some important meetings.	45%	60%
City Automotive Group (CAG) is a small company that developed out of a family-owned dealership that sells primarily American-manufactured cars. In addition to the sales function, there are four other key service components as well: Finance and Insurance (F&I), Body Shop , Parts & Accessories, and Service Department. Approximately five years ago, City Automotive Group (CAG was the #1 dealership in the country in terms of Customer Satisfaction for their Service Department. Since then, revenue has declined. The dealership owner has set a target of increasing the Service Department's annual revenue by $400K next year, and the department manager, Melvin Rogers, would like recommendations about how to attain this goal.	45%	50%

With nearly $90 billion in total assets, Smithsonian National is one of the largest financial and life insurance companies in the United States. Every year, Smithsonian's employees complete a satisfaction survey that measures their perceptions of the organization, their department, and the team they work on, as well as many other factors such as performance management practices, benefits, and compensation. After receiving last year's results, the Human Resources division is concerned about employees' negative perceptions about opportunities for career development within Smithsonian.	45%	50%
Management Solutions Consulting International (MSCI) is a publicly held business consulting and IT services firm. Their clients see their consultants as "information brokers". While individuals provide consulting services, clients expect that the collective skills within MSC are synergized and that experiences are applied rapidly to other projects. Unfortunately, MSCI consultants are geographically dispersed and there is a 23% annual turnover rate. Perhaps as a consequence, there is high turnover in MSCI's clients with little repeat business.	45%	40%
Expedit is a professional services organization that provides IT consulting services in a wide variety of industries. A major aspect of Expedit's business is application development and implementation, but many applications are being delivered late. This is leading to client dissatisfaction and millions of dollars in cost overrun.	45%	30%
Griffey, Horn, and Kensey is a large public accounting firm located in the Southwestern US. As with any accounting firm, auditors play a key role in the success of the organization. Unfortunately, resignations by newly hired auditors have increased over the past few years and last year's turnover for new auditors was almost 20%.	45%	30%
When performance problems occur at Evco, a utilities provider in Manfield County in southern Kentucky, they are typically solved with training programs. However, lately Evco has been underperforming on its financial indicators and as a result, cost-savings measures are being taken. As is often the case in these situations, the Training & Development program is being looked at as an area for possible cuts.	40%	44%

As an Education & Technology Director at a Department of Defense intelligence contractor, one of your priorities is language proficiency. There are three levels of language competency, but the minimum level to successfully perform missions is at the second level. Unfortunately, almost a third of the language analysts that you employ are below this minimum performance level.	40%	44%
Vive is a non-governmental organization operating in a community in Sub-Saharan Africa. Due to an epidemic of AIDs, almost 10,000 children have been orphaned in this community. Stretched by demands for service and a lack of funding, Vive's small staff of 7 depends on the few surviving women and teenagers in the community. These women and teenagers play the roles of caregiver and advocate for the orphaned children. Given the magnitude of the problem, they struggle with their own financial well being. Securing food and medication is a challenge and concerns that in the absence of sound sexual education courses, the community may never be able to recover from the AIDS epidemic.	40%	33%
PDRS is an online performance data reporting database with more than 1000 end users that are geographically distributed. Although PDRS was built by the internal IT department, there is no "help desk" support for the database. Instead, technical issues and troubleshooting is performed by a handful of staff in the department that owns the performance data contained in the system. These staff perform a variety of functions so they don't generally have a background in IT support and questions often require a week or more to resolve, which disrupts the end users' access to the performance data.	40%	22%
Orin is a large plumbing distributor in Ontario, Canada. Unfortunately, the warehouse staff are not familiar with plumbing parts, so they often select the wrong parts from inventory. As a result, there are many incorrect orders that are shipped. Incorrect shipments often result in complaints from clients, but there is also a cost implication as well. Sometimes higher cost parts are sent instead of lower cost items. In these cases, Orin depends on their customers' good will to report the issue.	40%	22%

Within a federal agency, the Program and Policy Development unit (PPDU) provides training, performance monitoring, and policy formulation services for approximately 80 field offices . PPDU has experienced a great deal of growth in the last few years, growing from a staff of six to 17. More recently, PPDU's manager has received several complaints from staff about how difficult it was to research questions concerning policy: staff are concerned the time that is wasted on searching for files and duplications of effort.	40%	22%
Statler is a major producer of information technology software, hardware, and gadgets. Having negotiated a contract with half a dozen school districts, Statler was outfitting classrooms with technology. As a part of these contracts, Statler was also providing training for teachers who would be using the equipment in the classroom. Unfortunately, there appears to be little change in teachers' usage of technology and no effect on student achievement.	40%	11%
In the mid 1990's, the US faced a banking crisis brought on by the greed and fraud of a few bank presidents. The lead agency charged with administering the national banking system, suspected that there were deficiencies in how banks were examined. Specifically, there was concern about in how the National Bank Examiners, NBEs, were prepared to do this work and that they lacked the requisite capabilities to identify deficiencies in loan portfolios.	36%	40%
Senior Support is an organization that encourages senior citizens' independence and involvement in the community. After conducting intake interviews with each senior citizen and the individuals interested in volunteering, the Volunteer Coordinators match several volunteers to a senior citizen, based on perceived personality match and volunteer availability. Unfortunately, the Senior Support's Director has received several angry or tearful calls from senior citizens who are unable to conduct their daily errands because their assigned volunteers are unavailable.	36%	20%

Tribute Awards is very small business that designs custom lapel pins, military coins, medallions, and other recognition products. Alongside increases in the prices of metals and changes in their off-shore manufacturer, profits have decreased. The owner, Marcus Coleman, has contacted you because he is having difficulty conveying policies and procedures to their employees. Ultimately, Marcus wants to improve customers' experience and increase the number of return customers.	30%	44%
Since the mid- 1980's, Del Mar Spas has been a successful international hot tub retail and manufacturing operation; in 2007 it's sales peaked at $52 million. But since the US housing market crisis, Del Mar's sales have steadily declined to almost $24 million a year. Del Mar's Founder, Jim Haddock, has received bids to buy the company, but deemed them all too low. In the face of dwindling sales and a dramatic change in the market, Jim has contacted you to help him find a way to help the company survive, and hopefully, thrive in the future.	30%	22%
For some time, student achievement at Martin Luther King High School has been low. In addition to low achievement on standardized tests, graduation rates have steadily declined to near 25%. Moreover, as most of the students come from families of low socio-economic status, very few students have access to a computer at home or plan to attend college.	30%	22%
Beaubien is an elementary education (K-8) school in the Midwest. It is part of a struggling school district, located in Braxton Park. Once a manufacturing hub, Braxton Park now battles high crime rates and is laboring to survive—it has been in receivership for several years. In the midst of this, Beaubien's teachers and staff aim to have an impact on students beyond the classroom into a self-sufficient adult life. However, Beaubien's students are not performing well on standardized tests and the school is not making adequate yearly progress on test scores. Because of the No Child Left Behind act, there are several negative consequences that are possible for Beaubien. Concerned about this prospect, the principal, Dr. Samson Jones, has contacted you.	30%	22%
cases are seen in the ER instead of primary or urgent care facilities, there are unnecessary costs. Additionally, human resources are diverted from patients with more serious needs.	25%	14%

Zen Tea Co has several small distribution centers and production plants across the country. As the director of HR at one of these plants, you have recently received a the production supervisors to revise policies for breaks. Apparently, production operators have been taking too many breaks. Currently those operators who perform tedious procedures received several extra 5 minute breaks a day. Often, these breaks ended up being longer than five minutes, with all operators (not just those doing tedious procedures) feeling they were also entitled to the extra breaks. As a result, there are conflicts between work groups and production is behind in its output. The supervisors also are frustrated because scheduling breaks takes away from their other responsibilities.	20%	11%
RightFast is small tax preparation company. The company started just a year ago when the owner, decided to retire from her position as a tax preparer at a larger tax prep firm. Due to the long working relationships, she decided to open her own firm, since many of her clients would follow her. Now in its second year, RightFast also employs 5 other seasonal employees. Despite previous performance and the company's name, many extensions were being filed and clients' tax returns were being completed late.	20%	11%
Janssen Smith is the CIO at an R-I university. Recently the Provost has mandated that the University's business processes be transformed. Furthermore, these new processes should "leverage enterprise systems technologies to include highly valuable features, such as workflow management, imaging, and advanced reporting". Mr. Smith is especially focused reforming the Information Technology Help Desk, due to a high volume of complaints from University staff about the number of calls required in order to resolve a problem.	18%	20%

Click to write Statement 10Professor Ames teaches graduate level courses in a business technology-oriented program. The field as a whole is relatively new: it began in the early 1950's but began to grow exponentially in the 1970's. Perhaps due to the newness of the field, there are few comprehensive texts for novice practitioners in the field. For those that do, students with more professional experience do well with them. But, Dr. Ames and her colleagues have noticed that novices struggle with the content of these texts. In addition to asking more questions, taking longer to master the content, course evaluations support that students who are inexperienced in the field have less satisfaction with the existing textbooks.	10%	0%
Medit is a second generation family-owned business that designs, develops, and manufactures highly specialized medical devices. Despite the complexity of it's business, Medit's senior leaders have vast responsibility for day-to-day functions. The Sales and Marketing VP retired last year and the owner, Bryce Stockton, is concerned that the position is still not filled.	9%	30%

APPENDIX E – RANK ORDERING OF TYPES OF EVIDENCE BY EXPERT PANEL RATINGS

Question	Ratio of Science to Craft Ratings
Article in research-oriented journal	n/a
Content analysis findings	n/a
Research journals in the field of education	10
Research journals in the field of business	10
Research journals in the field of psychology	10
Findings of prescriptive research	9
Graphical representation of cause-effect relationships	9
Findings from descriptive research	8
Learning theories	4
Examining similar interventions that have been designed to meet the same objectives	3.5
Logical reasoning	2.5
Observing others' use of an intervention	2.333333333
Tests of a prototype	2.333333333
Reason-checking with previous work experiences	2
PIJ articles	1.75
Reason-checking with previous life experiences	1.666666667
Consultation with subject matter experts	1.5
Informal conversations to reason-check thinking	1.5
Presentations at AECT conferences	1.333333333
Client interviews	1.333333333
Prescriptive theories about intervention design	1.25
Stakeholder feedback	1.25
University-sponsored internet forums	1.166666667
Practitioner journals in the field of psychology	1.142857143
Subject matter expert interviews	1.142857143
Domain-based principles of learning	1
Guidelines for appropriate solutions based on a class of causes	1
Project team knowledge	1
Textbooks	1
Presentations at ISPI conferences	1
Trial-and-error	0.888888889

Books outside the field of PI	0.8
Discussion with a trusted colleague	0.8
Formal reflection on alternatives in writing	0.8
Hypothetical target audience reactions	0.8
Initial information provided by client	0.8
Prescriptive instructional design theories	0.8
Practitioner journals in the field of business	0.75
Client provided descriptions of the performance problem	0.75
Article in a practitioner-oriented journal	0.666666667
Discussion with the client	0.666666667
Visualizations of intervention results	0.666666667
Subject matter expert opinions	0.666666667
Other internal PI professionals' opinions	0.6
An in-house training course on intervention selection	0.6
Presentations at ASTD conferences	0.555555556
Feedback from a colleague who has an opposing style	0.5
Generalizations from everyday experience	0.5
Ideas	0.5
Journaling personal reflections	0.5
Practitioner journals in the field of education	0.5
Talking with someone else who has implemented the intervention	0.5
Client opinions	0.5
Memories of previous work experiences	0.5
Memories of previous experience as a student	0.5
Memories of previous experience as an instructor	0.5
A training course on intervention selection developed by a vendor	0.5
Article in a professional magazine	0.428571429
Consultation with target audience regarding their preferences	0.428571429
Discussion with peers or colleagues	0.428571429
Envisioning the results of various possible scenarios	0.428571429
Reflection on conventional wisdom	0.428571429
Gut feelings	0.363636364
Intuitions	0.363636364
A sub-conscious breakthrough while doing something else	0.285714286
Day dreaming	0.285714286
Social-networking internet forums	0.272727273
Alumni listservs	0.272727273
A template for intervention created by someone else	0.25
A template for intervention selection created by yourself	0.25

Brainstorming individually	0.25
Brainstorming with another colleague on the project	0.25
Description of a specific project that doesn't draw any conclusions	0.25
Descriptions of popular practices	0.25
Internet websites	0.25
Personally constructed mental models	0.25
Reflections on subconscious	0.25
Trade books	0.25
Creative flashes	0.125
Dreams about the situation	0.125
Hunches	0.125
Instincts	0.125
Clients' preferences	0.111111111
Flash of genius	0.111111111
Perceptions of the performance project team's values	0.111111111
Personal insight	0.111111111
Personal theories of practice	0.111111111
Recalling a similar situation from past experience	0.111111111
Stroke of insight	0.111111111
Commentary article	0
Description of a specific project that doesn't include data	0
Editorial article	0
Hopes	0
Imagination	0
Personal feelings	0
Personal values	0

Note: N/A indicates situations where all of the experts marked that the type of evidence was science-based.

APPENDIX F – RESPONSIVENESS RESULTS FOR PILOT STUDY

Scenario Name	Evidence Description	*2l*	*l*	*p*
SS1	Findings of prescriptive research	0.702	0.351	0.402
SS8	Case study	5.188	2.594	0.023
SS10*	Staff interviews	0.29	0.145	0.59
SS16	Graphical representation of cause and effect	6.416	3.208	0.011
AD2	Hunch	-19.95	-9.975	1
AD5	Client interview	-19.836	-9.918	1
AD9*	Personal insight	-11.94	-5.97	1
AD15	Brainstorming with a client	-31.102	-15.551	1
SD3*	Multiple research studies that fail to reject null hypothesis	-10.28	-5.14	1
SD6	Subject matter expert consultation	-32.8	-16.4	1
SD11	Examining results of implementation of intervention designed to meet similar objectives	-26.46	-13.23	1
SD13	Pilot testing a prototype	-33.6	-16.8	1
AS4*	Dreaming	2.734	1.367	0.09
AS7	Internet research	4.766	2.383	0.029
AS12	Memories of past experience	7.388	3.694	0.006
AS14	Editorial article	3.382	1.691	0.065

Note. Scenarios denoted with an asterisk were eliminated from the final instrument.

APPENDIX G–EMAIL INVITATION & REMINDER

Dear [First Name],
You are invited you to participate in an online questionnaire of intervention selection decisions. It will take you approximately 20 -25 minutes to complete this questionnaire. Participants will have the option of entering a random drawing to receive one of fifty $25.00 restaurant.com gift certificates.

Your participation is crucial to the findings of this study and the future of our field as a technology.

Participation is as easy as clicking this link

https://acsurvey.qualtrics.com/SE/?SID=SV_0BWMUOXcsXOgq21

If you choose to participate in the drawing, you will need to provide your e-mail address. Your email address will only be used to administer the incentive.
Please e-mail or call me if you have questions on participating in or learning more about this study. I can be reached at Name@gmail.com or please call (###) ###-####.

Please do not participate if you have already taken part in earlier phases of this study.
Dear [FirstName],

There is still time to participate in an online questionnaire of intervention selection decisions. It will take you approximately 20 -25 minutes to complete this questionnaire. Participants will have the option of entering a random drawing to receive one of fifty $25.00 restaurant.com gift certificates.

Your participation is crucial to the findings of this study and the future of our field as a technology.

Participation is as easy as clicking this link

https://acsurvey.qualtrics.com/SE/?SID=SV_0BWMUOXcsXOgq21

If you choose to participate in the drawing, you will need to provide your e-mail address. Your email address will only be used to administer the incentive.
Please e-mail or call me if you have questions on participating in or learning more about this study. I can be reached at Name@gmail.com or please call (###) ###-####.

Please do not participate if you have already taken part in earlier phases of this study.

APPENDIX H – PRIOR DISTRIBUTION FOR RQ6

Analysis for RQ6a

The CORR Procedure

Intervention_type=1

2 Variables:	Familiarity P1

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	233	4.17597	0.87046	4.00000	2.00000	5.00000	Familiarity
P1	233	71.54936	18.63114	75.00000	11.00000	100.00000	P1

Spearman Correlation Coefficients, N = 233
Prob > |r| under H0: Rho=0

	Familiarity	P1
Familiarity Familiarity	1.00000	0.34578 <.0001
P1 P1	0.34578 <.0001	1.00000

Spearman Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	**P1**	233	0.34578	0.36065	0.0007452	0.34513	0.226660	0.453532	<.0001

Analysis for RQ6a

The CORR Procedure

Intervention_type=2

2 Variables:	Familiarity P1

Simple Statistics							
Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	224	4.17411	0.92840	4.00000	2.00000	5.00000	Familiarity
P1	227	71.58150	20.97251	80.00000	20.00000	100.00000	P1

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	P1
Familiarity Familiarity	1.00000 224	0.15023 0.0245 224
P1 P1	0.15023 0.0245 224	1.00000 227

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	P1	224	0.15023	0.15137	0.0003368	0.14990	0.019191	0.275565	0.0244

Analysis for RQ6a

The CORR Procedure

Intervention_type=3

2 Variables:	Familiarity P1

Simple Statistics							
Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	35	3.85714	1.00419	4.00000	2.00000	5.00000	Familiarity
P1	36	63.83333	22.86544	63.00000	9.00000	95.00000	P1

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	P1
Familiarity Familiarity	1.00000 35	0.20228 0.2439 35
P1 P1	0.20228 0.2439 35	1.00000 36

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	P1	35	0.20228	0.20511	0.00297	0.19943	-0.143348	0.499478	0.2459

Analysis for RQ6a

The CORR Procedure

Intervention_type=4

2 Variables:	Familiarity P1

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	169	4.59172	0.92848	5.00000	1.00000	5.00000	Familiarity
P1	169	63.61538	23.20099	64.00000	12.00000	100.00000	P1

Spearman Correlation Coefficients, N = 169 Prob > |r| under H0: Rho=0		
	Familiarity	**P1**
Familiarity Familiarity	1.00000	0.23311 0.0023
P1 P1	0.23311 0.0023	1.00000

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P1**	169	0.23311	0.23747	0.0006938	0.23245	0.084455	0.370414	0.0022

Analysis for RQ6a

The CORR Procedure

Intervention_type=5

2 Variables:	Familiarity P1

Simple Statistics							
Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	58	3.43103	1.25813	3.00000	1.00000	5.00000	Familiarity
P1	57	52.38596	20.95553	50.00000	3.00000	95.00000	P1

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	P1
Familiarity Familiarity	1.00000 58	0.57626 <.0001 57
P1 P1	0.57626 <.0001 57	1.00000 57

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	P1	57	0.57626	0.65685	0.00515	0.57282	0.367031	0.725150	<.0001

Analysis for RQ6a

The CORR Procedure

Intervention_type=6

2 Variables:	Familiarity P1

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	56	3.42857	1.09307	3.50000	2.00000	5.00000	Familiarity
P1	57	66.38596	21.68175	70.00000	2.00000	94.00000	P1

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	**P1**
Familiarity Familiarity	1.00000 56	0.18480 0.1727 56
P1 P1	0.18480 0.1727 56	1.00000 57

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P1**	56	0.18480	0.18695	0.00168	0.18318	-0.083755	0.425584	0.1735

Analysis for RQ6b

The CORR Procedure

Intervention_type=1

2 Variables:	Familiarity P2

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	233	4.17597	0.87046	4.00000	2.00000	5.00000	Familiarity
P2	233	72.53219	16.97564	75.00000	20.00000	100.00000	P2

Spearman Correlation Coefficients, N = 233 Prob > |r| under H0: Rho=0		
	Familiarity	**P2**
Familiarity Familiarity	1.00000	0.13692 0.0367
P2 P2	0.13692 0.0367	1.00000

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P2**	233	0.13692	0.13778	0.0002951	0.13663	0.008253	0.260575	0.0367

Analysis for RQ6b

The CORR Procedure

Intervention_type=2

2 Variables:	Familiarity P2

Simple Statistics							
Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	224	4.17411	0.92840	4.00000	2.00000	5.00000	Familiarity
P2	227	73.02203	16.77346	75.00000	29.00000	100.00000	P2

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	P2
Familiarity Familiarity	1.00000 224	0.23351 0.0004 224
P2 P2	0.23351 0.0004 224	1.00000 227

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	P2	224	0.23351	0.23790	0.0005236	0.23301	0.105141	0.353304	0.0004

Analysis for RQ6b

The CORR Procedure

Intervention_type=3

2 Variables:	Familiarity P2

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	35	3.85714	1.00419	4.00000	2.00000	5.00000	Familiarity
P2	36	70.55556	16.89369	70.00000	30.00000	100.00000	P2

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	**P2**
Familiarity Familiarity	1.00000 35	0.57757 0.0003 35
P2 P2	0.57757 0.0003 35	1.00000 36

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P2**	35	0.57757	0.65881	0.00849	0.57188	0.294825	0.760245	0.0002

Analysis for RQ6b

The CORR Procedure

Intervention_type=4

2 Variables:	Familiarity P2

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	169	4.59172	0.92848	5.00000	1.00000	5.00000	Familiarity
P2	169	74.74556	15.91309	79.00000	25.00000	100.00000	P2

Spearman Correlation Coefficients, N = 169
Prob > |r| under H0: Rho=0

	Familiarity	P2
Familiarity Familiarity	1.00000	0.18667 0.0151
P2 P2	0.18667 0.0151	1.00000

Spearman Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	**P2**	169	0.18667	0.18889	0.0005556	0.18614	0.036194	0.327884	0.0149

Analysis for RQ6b

The CORR Procedure

Intervention_type=5

2 Variables:	Familiarity P2

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	58	3.43103	1.25813	3.00000	1.00000	5.00000	Familiarity
P2	58	66.84483	14.98163	70.00000	30.00000	91.00000	P2

Spearman Correlation Coefficients, N = 58 Prob > |r| under H0: Rho=0		
	Familiarity	**P2**
Familiarity Familiarity	1.00000	0.09623 0.4724
P2 P2	0.09623 0.4724	1.00000

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P2**	58	0.09623	0.09653	0.0008441	0.09539	-0.167018	0.345184	0.4741

Analysis for RQ6b

The CORR Procedure

Intervention_type=6

2 Variables:	Familiarity P2

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	56	3.42857	1.09307	3.50000	2.00000	5.00000	Familiarity
P2	57	64.98246	16.51243	64.00000	14.00000	90.00000	P2

Spearman Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Familiarity	P2
Familiarity Familiarity	1.00000 56	0.10956 0.4215 56
P2 P2	0.10956 0.4215 56	1.00000 57

Spearman Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	**P2**	56	0.10956	0.11001	0.0009960	0.10858	-0.158855	0.361171	0.4232

Analysis for RQ6c

The CORR Procedure

Intervention_type=1

2 Variables:	Familiarity P3

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	233	4.17597	0.87046	4.00000	2.00000	5.00000	Familiarity
P3	225	58.50667	26.72548	60.00000	0	100.00000	P3

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	**P3**
Familiarity Familiarity	1.00000 233	0.09718 0.1462 225
P3 P3	0.09718 0.1462 225	1.00000 225

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P3**	225	0.09718	0.09749	0.0002169	0.09696	-0.034262	0.224902	0.1464

Analysis for RQ6c

The CORR Procedure

Intervention_type=2

2 Variables:	Familiarity P3

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	224	4.17411	0.92840	4.00000	2.00000	5.00000	Familiarity
P3	222	64.45495	26.11210	70.00000	0	100.00000	P3

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	**P3**
Familiarity Familiarity	1.00000 224	0.07932 0.2424 219
P3 P3	0.07932 0.2424 219	1.00000 222

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P3**	219	0.07932	0.07949	0.0001819	0.07914	-0.053998	0.209518	0.2427

Analysis for RQ6c

The CORR Procedure

Intervention_type=3

2 Variables:	Familiarity P3

Simple Statistics							
Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	35	3.85714	1.00419	4.00000	2.00000	5.00000	Familiarity
P3	34	60.88235	25.29554	66.50000	3.00000	100.00000	P3

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	P3
Familiarity Familiarity	1.00000 35	0.48161 0.0045 33
P3 P3	0.48161 0.0045 33	1.00000 34

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	P3	33	0.48161	0.52508	0.00753	0.47581	0.158370	0.704103	0.0040

Analysis for RQ6c

The CORR Procedure

Intervention_type=4

2 Variables:	Familiarity P3

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	169	4.59172	0.92848	5.00000	1.00000	5.00000	Familiarity
P3	165	70.80000	24.47278	80.00000	0	100.00000	P3

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	**P3**
Familiarity Familiarity	1.00000 169	0.14925 0.0557 165
P3 P3	0.14925 0.0557 165	1.00000 165

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P3**	165	0.14925	0.15037	0.0004550	0.14881	-0.004071	0.294885	0.0556

Analysis for RQ6c

The CORR Procedure

Intervention_type=5

2 Variables:	Familiarity P3

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**	**Label**
Familiarity	58	3.43103	1.25813	3.00000	1.00000	5.00000	Familiarity
P3	58	55.62069	29.68302	50.50000	2.00000	100.00000	P3

Spearman Correlation Coefficients, N = 58 Prob > |r| under H0: Rho=0		
	Familiarity	**P3**
Familiarity Familiarity	1.00000	0.20394 0.1247
P3 P3	0.20394 0.1247	1.00000

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	**With Variable**	**N**	**Sample Correlation**	**Fisher's z**	**Bias Adjustment**	**Correlation Estimate**	**95% Confidence Limits**		**p Value for H0:Rho=0**
Familiarity	**P3**	58	0.20394	0.20684	0.00179	0.20222	-0.059162	0.437659	0.1250

Analysis for RQ6c

The CORR Procedure

Intervention_type=6

2 Variables:	Familiarity P3

Simple Statistics							
Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
Familiarity	56	3.42857	1.09307	3.50000	2.00000	5.00000	Familiarity
P3	54	55.55556	28.63279	55.00000	0	100.00000	P3

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Familiarity	P3
Familiarity Familiarity	1.00000 56	0.13861 0.3222 53
P3 P3	0.13861 0.3222 53	1.00000 54

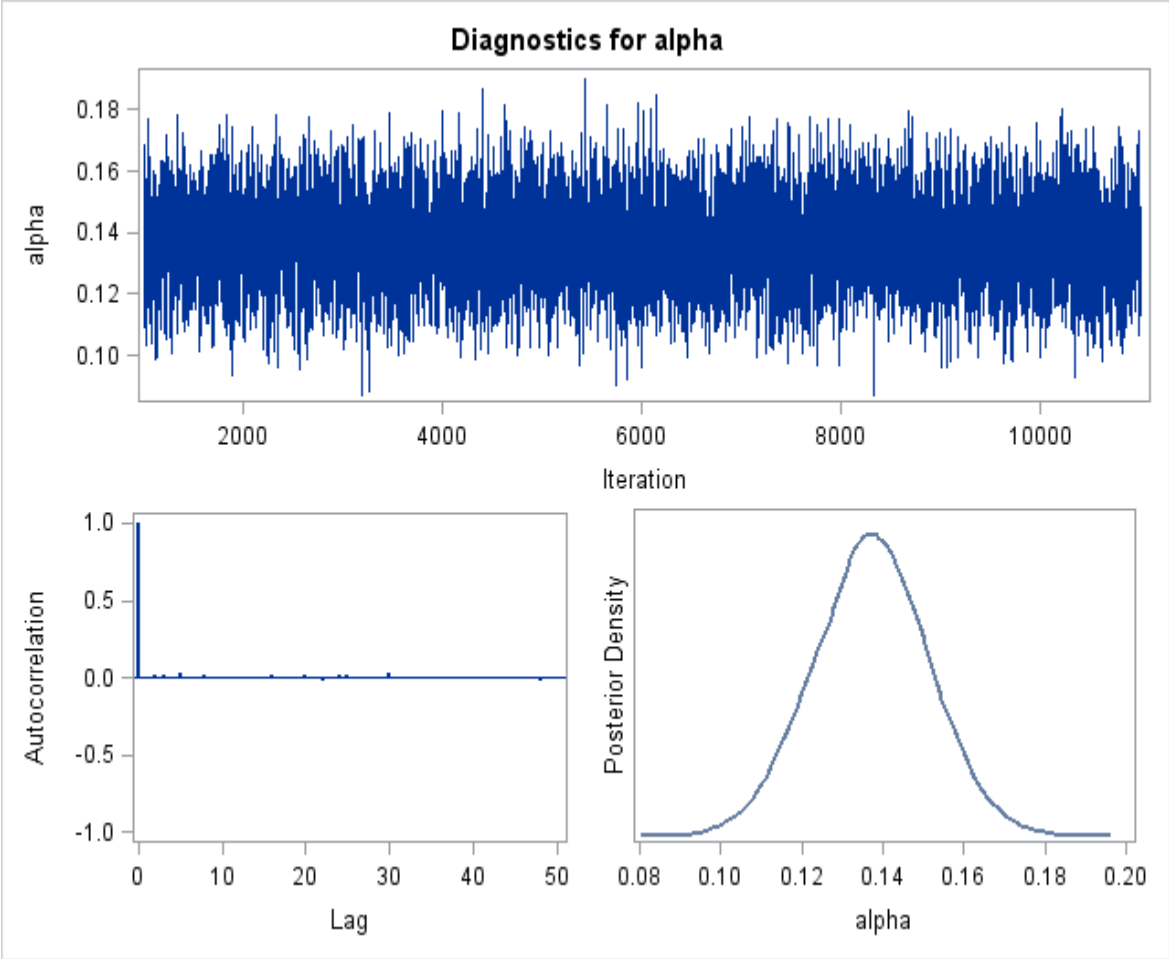
Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
Familiarity	P3	53	0.13861	0.13951	0.00133	0.13731	-0.138113	0.393014	0.3239

APPENDIX I – DENSITY PLOTS FOR RQ6

General Assessments - Pr_1

Intervention Type: Data

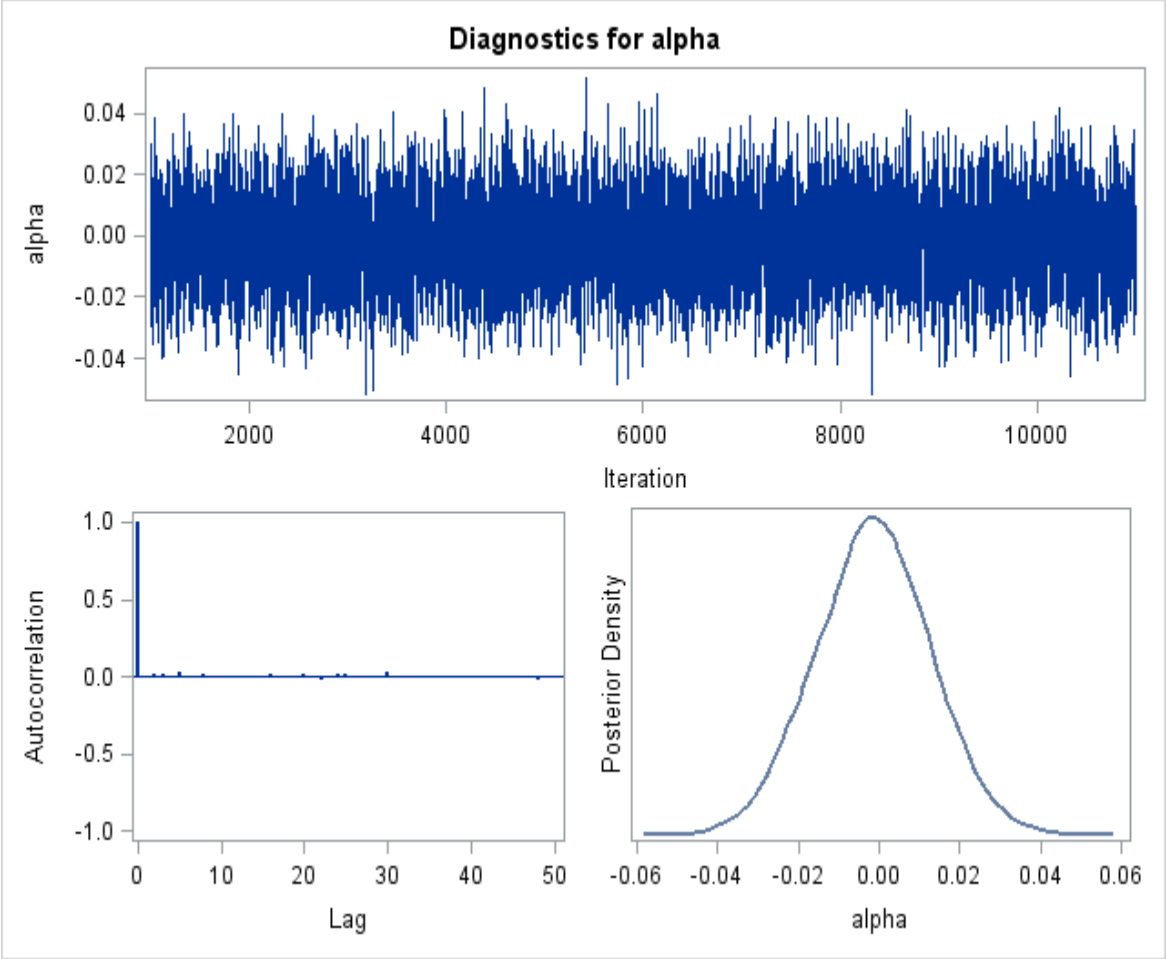
The MCMC Procedure



General Assessments - Pr_1

Intervention Type: Instrumentation

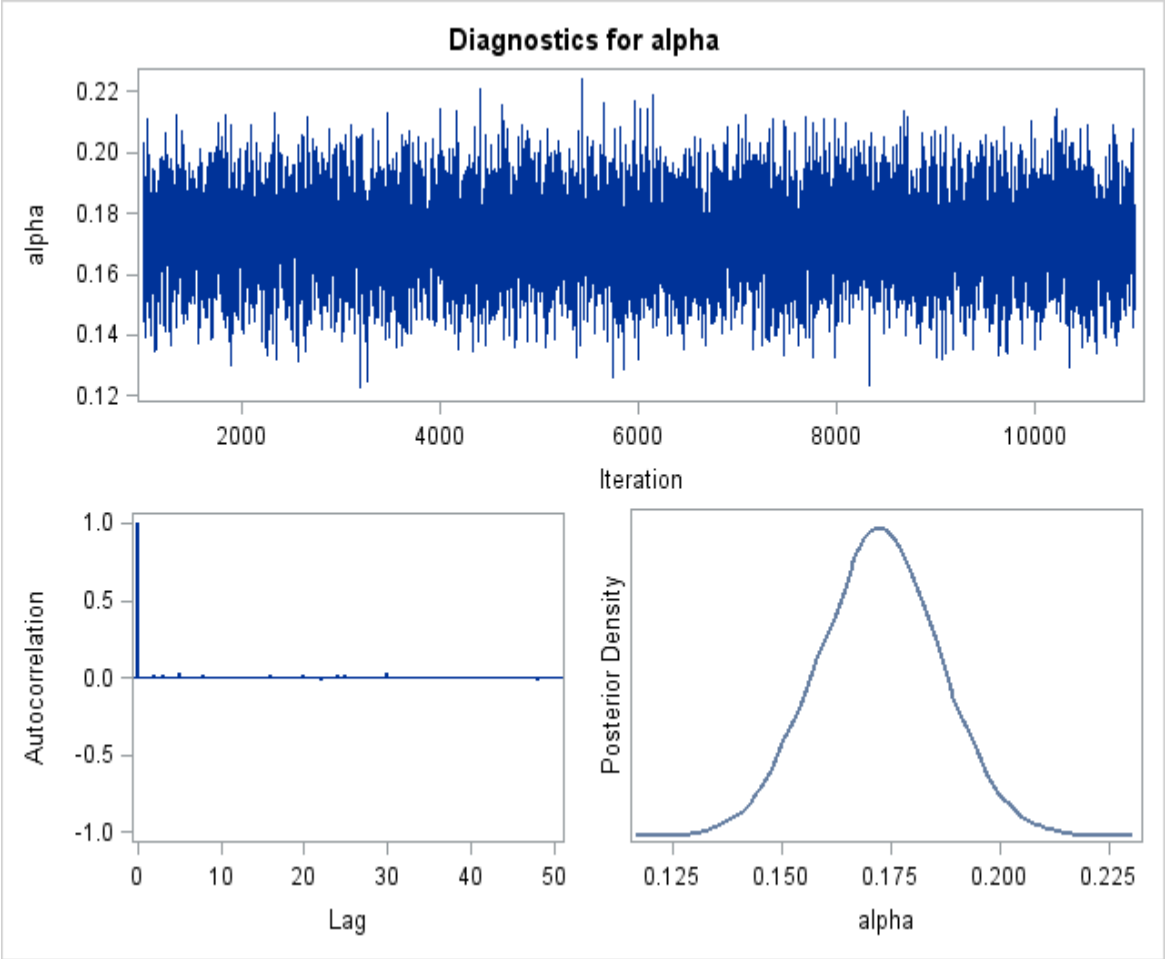
The MCMC Procedure



General Assessments - Pr_1

Intervention Type: Incentives

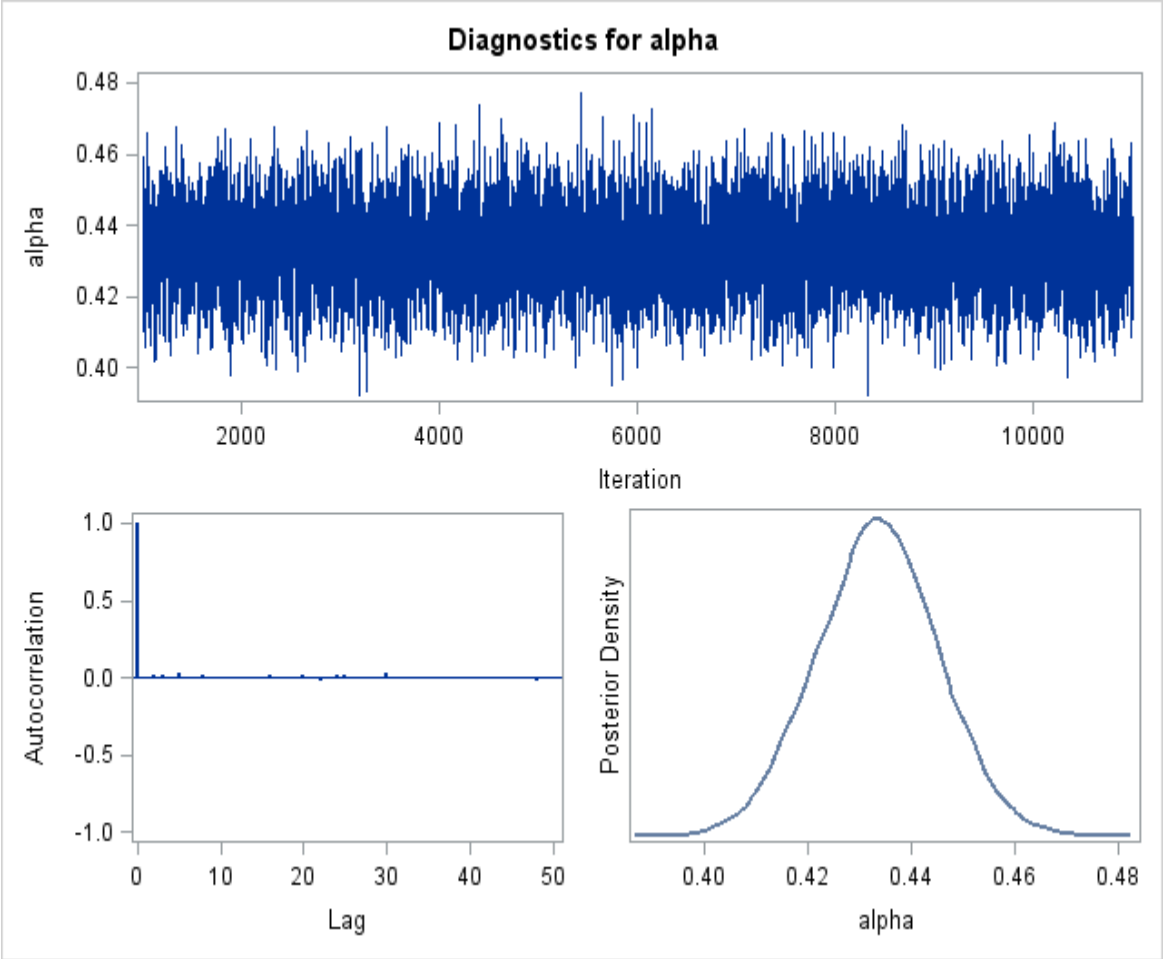
The MCMC Procedure



General Assessments - Pr_1

Intervention Type: Knowledge

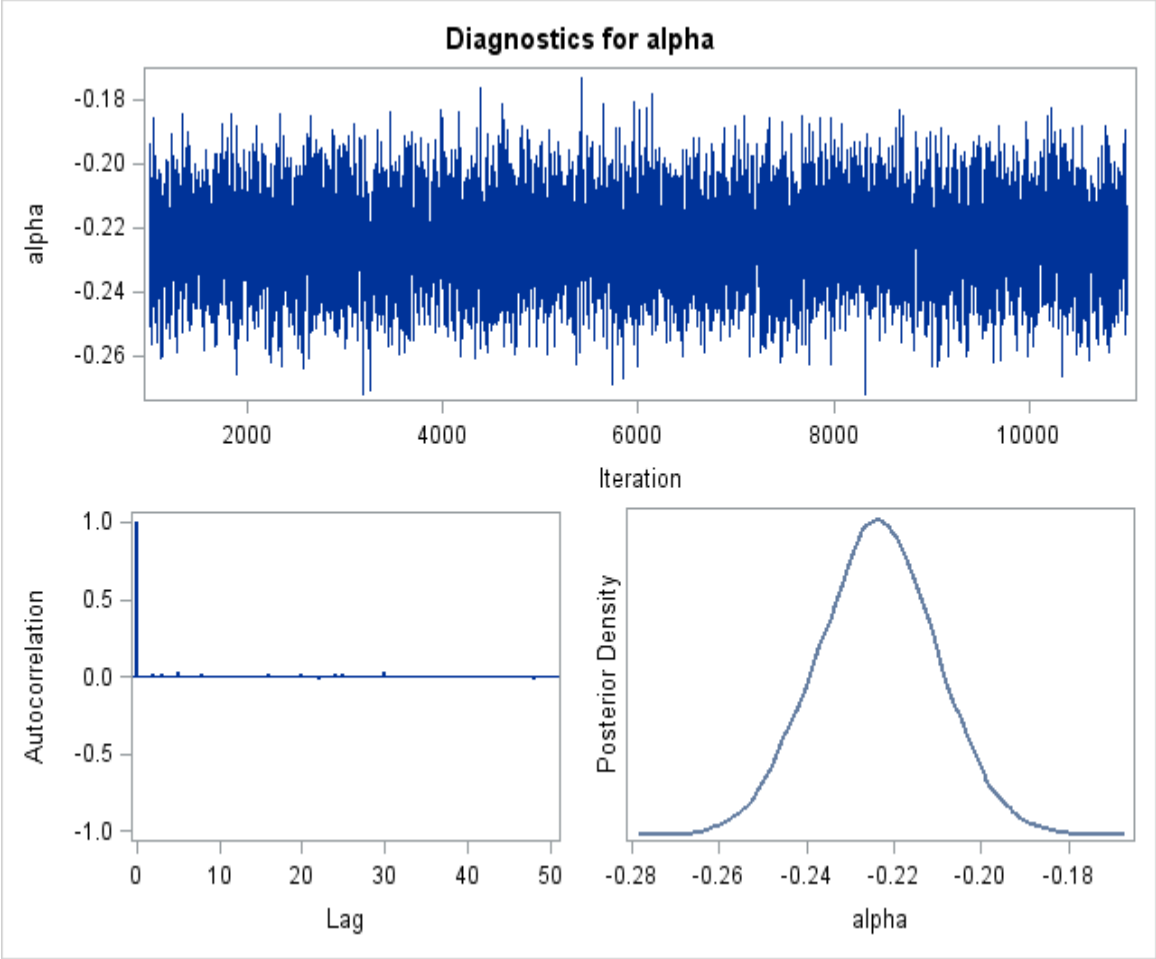
The MCMC Procedure



General Assessments - Pr_1

Intervention Type: Capacity

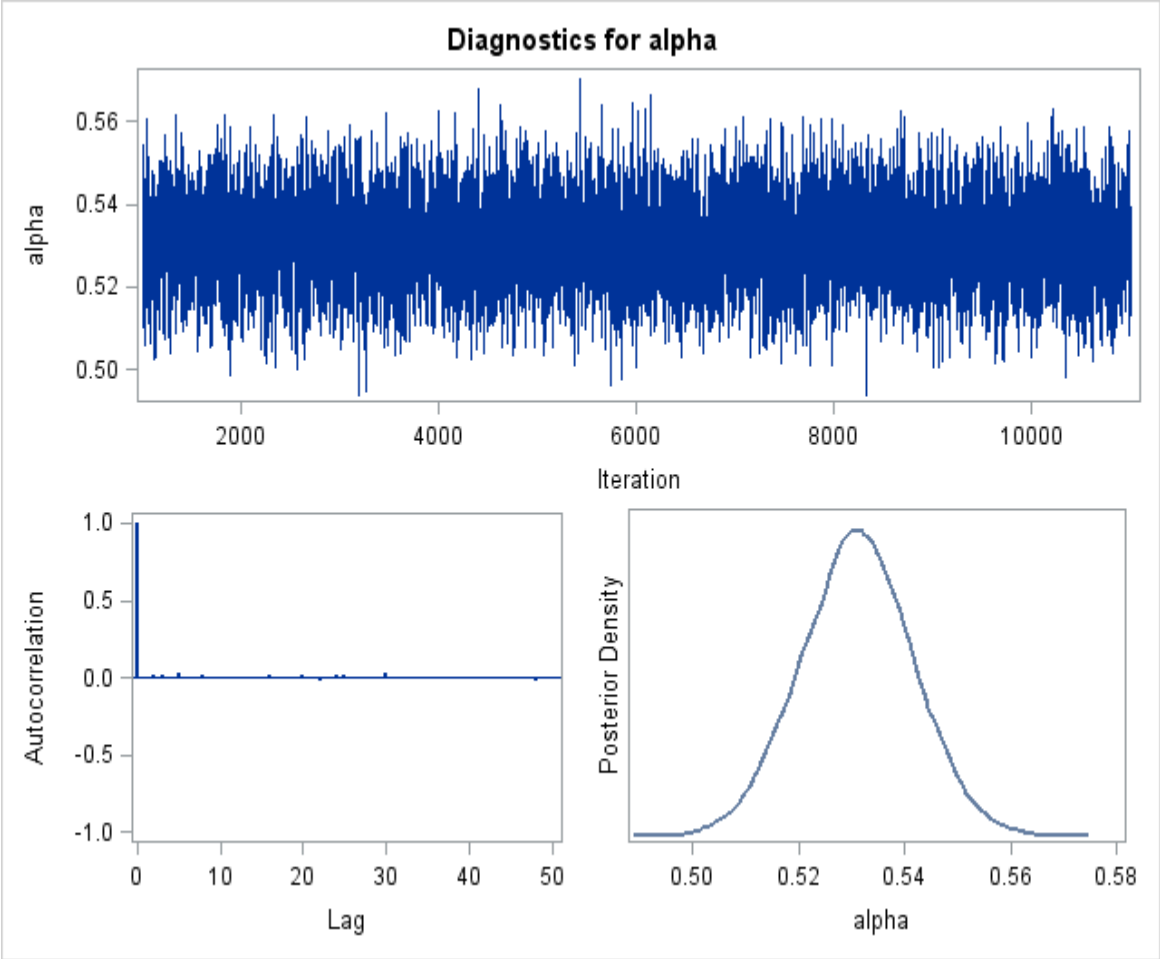
The MCMC Procedure



General Assessments - Pr_1

Intervention Type: Motives

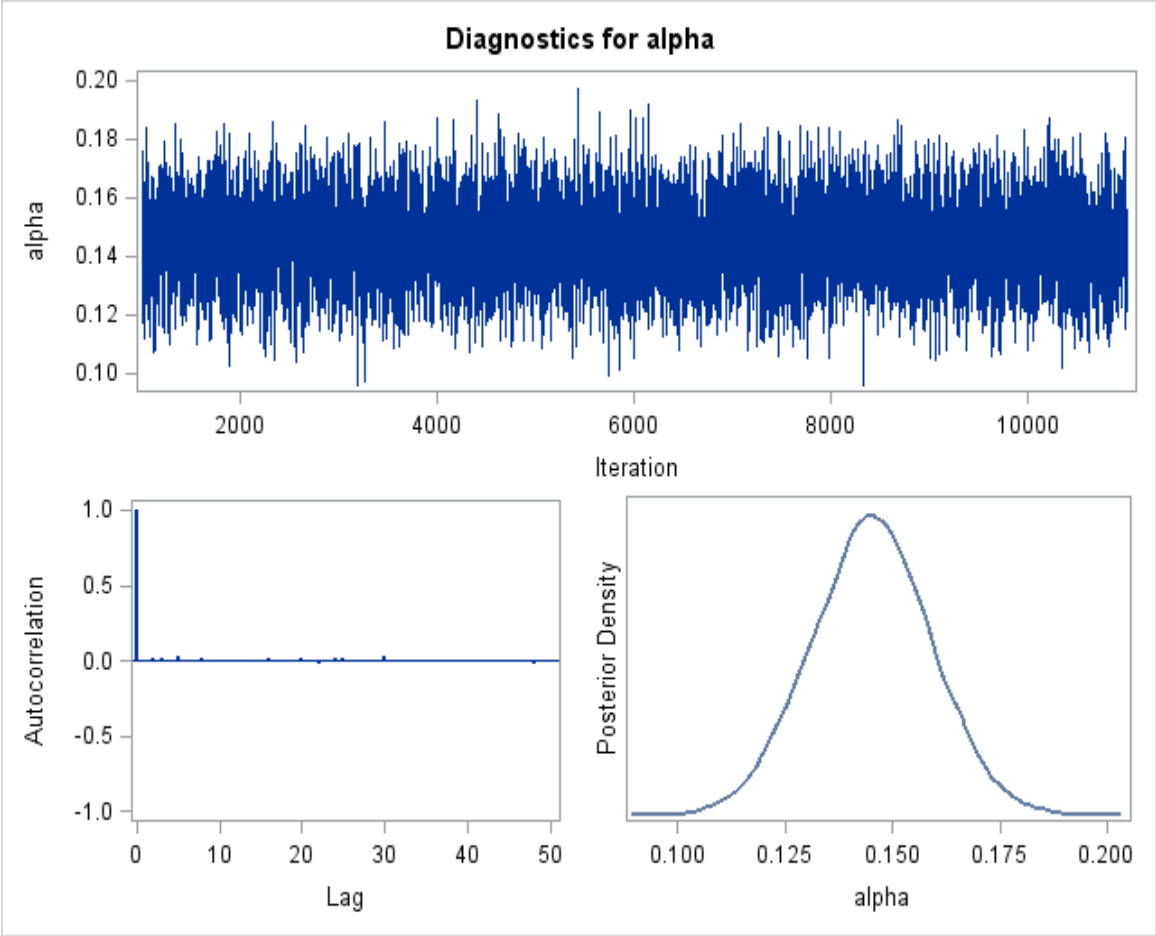
The MCMC Procedure



Priors – Pr_2

Intervention Type: Data

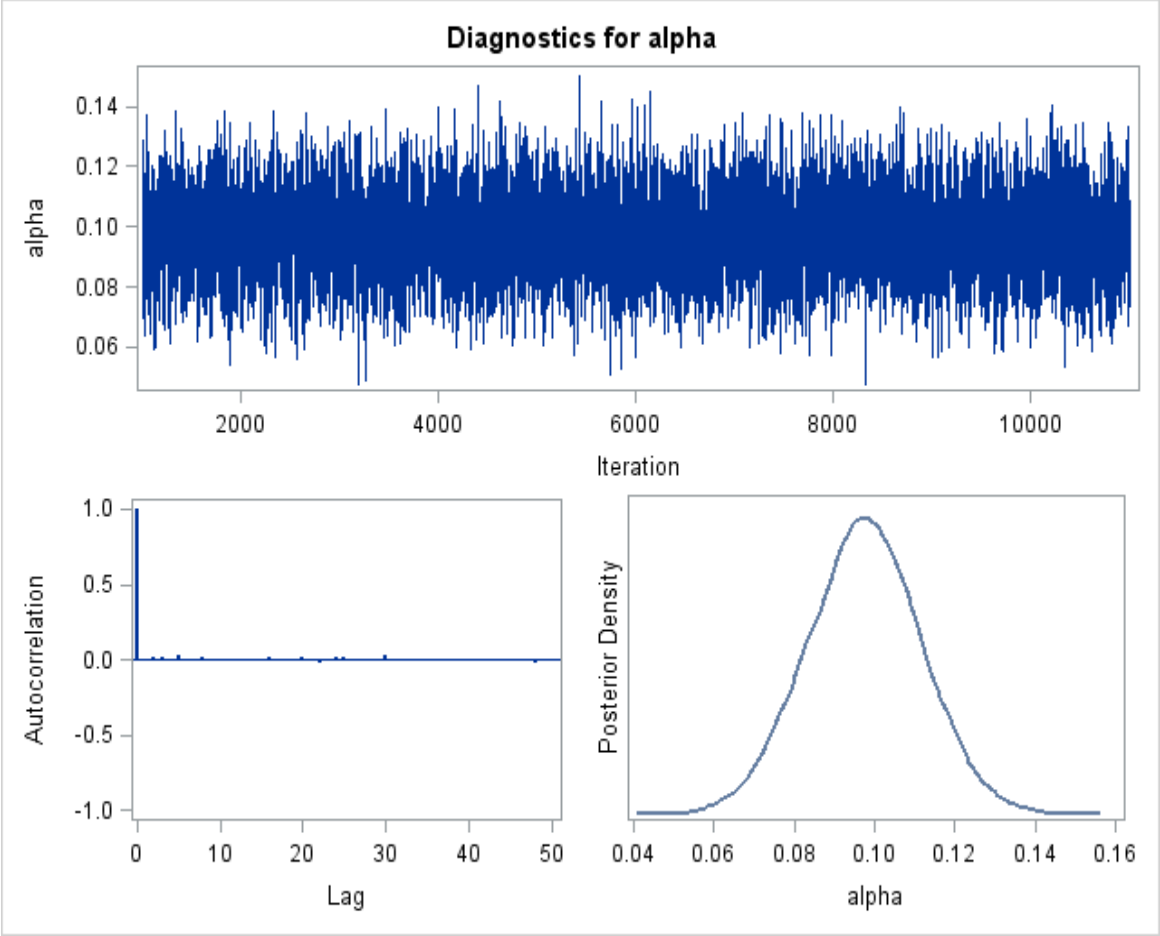
The MCMC Procedure



Priors – Pr_2

Intervention Type: Instrumentation

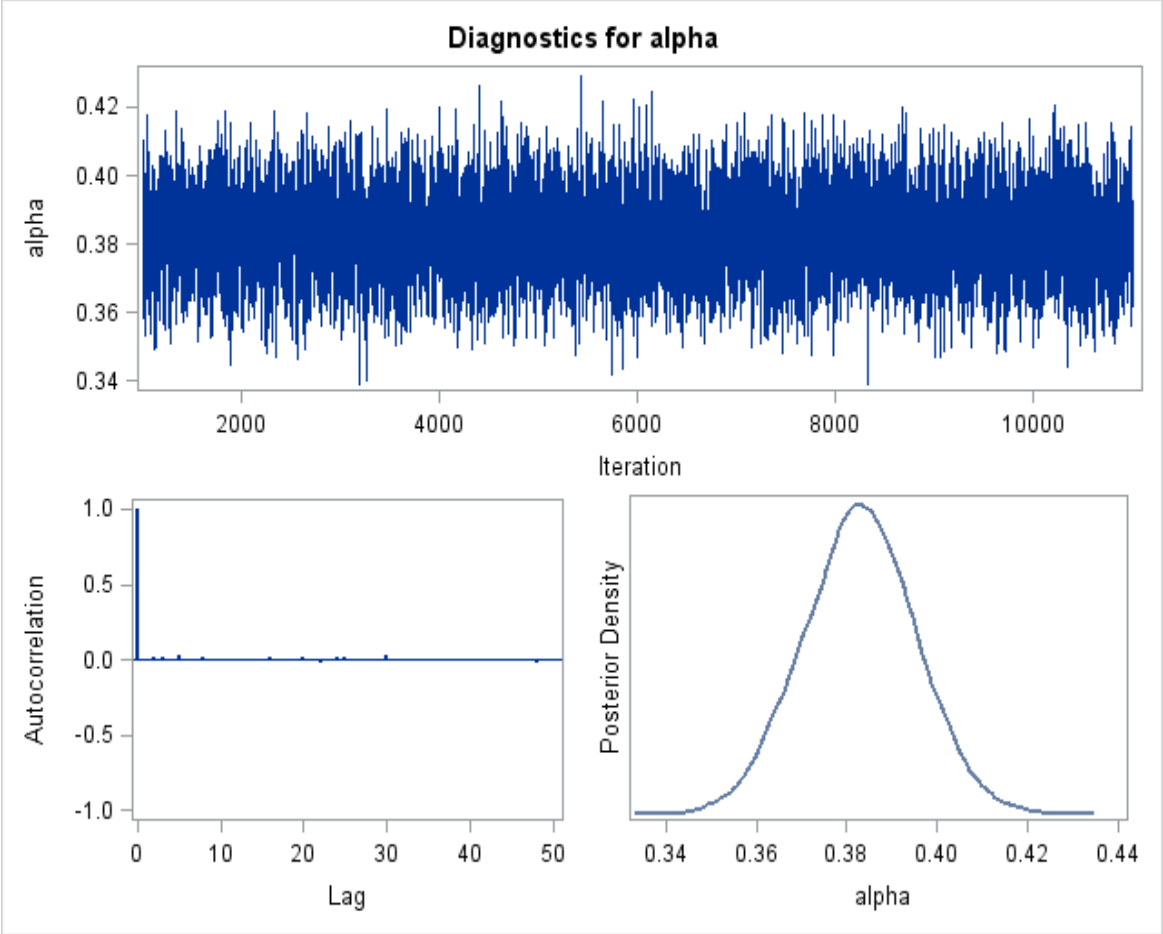
The MCMC Procedure



Priors – Pr_2

Intervention Type: Incentives

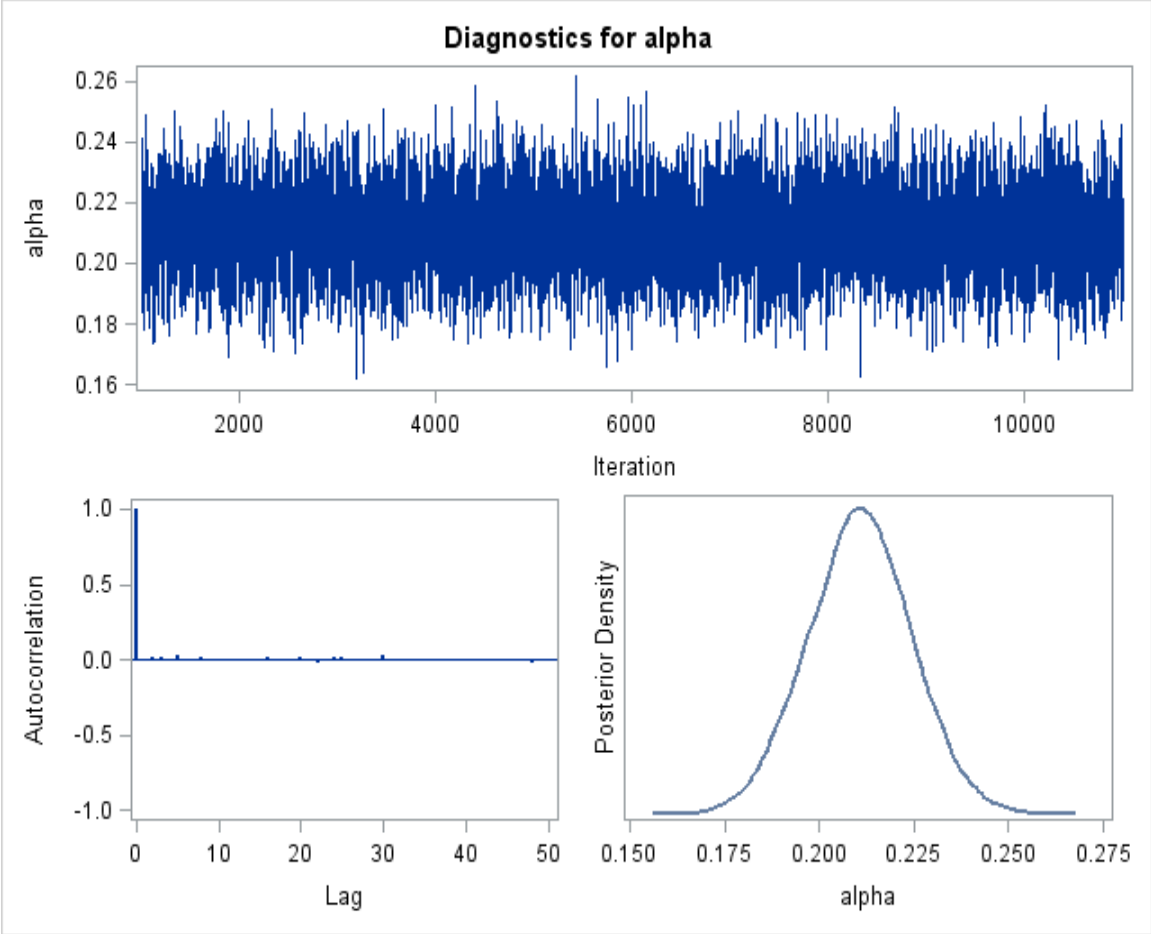
The MCMC Procedure



Priors – Pr_2

Intervention Type: Knowledge

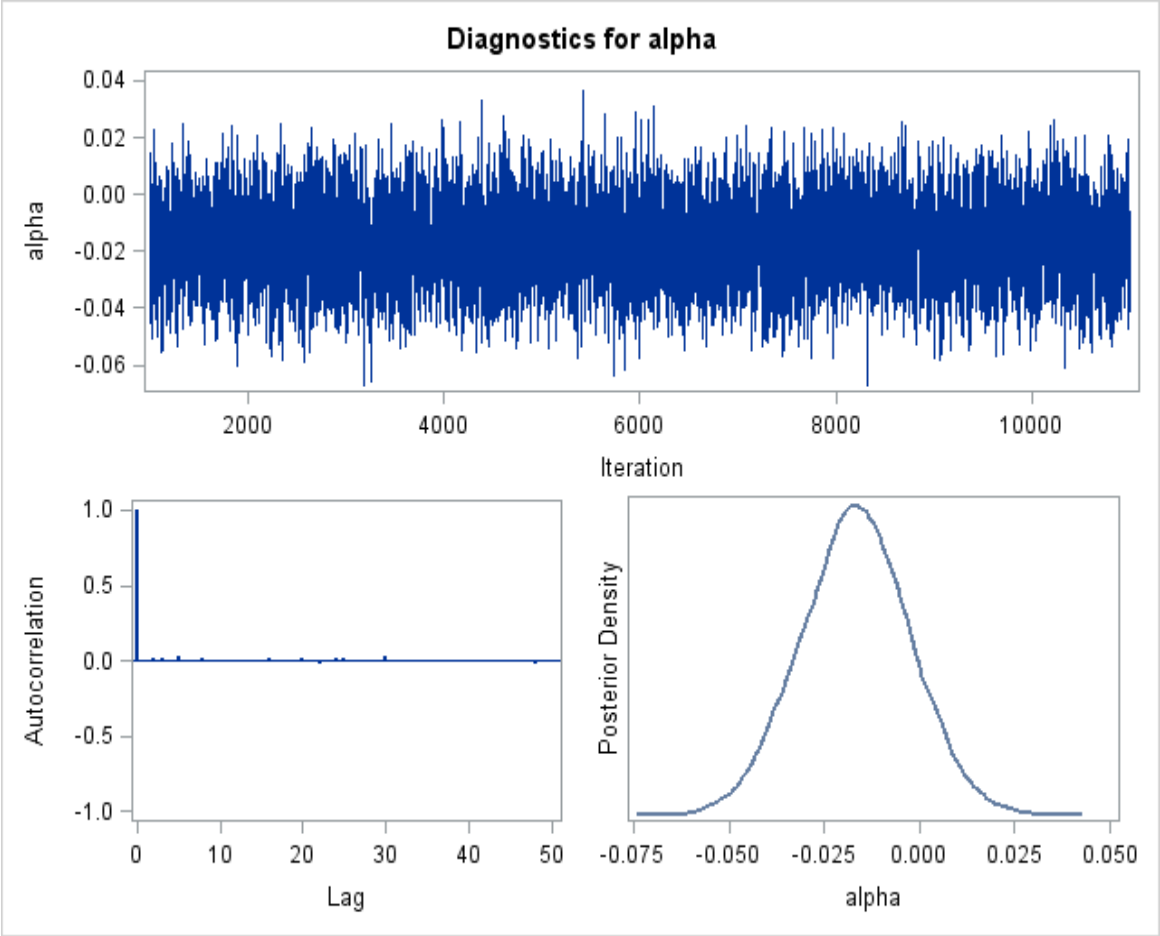
The MCMC Procedure



Priors – Pr_2

Intervention Type: Capacity

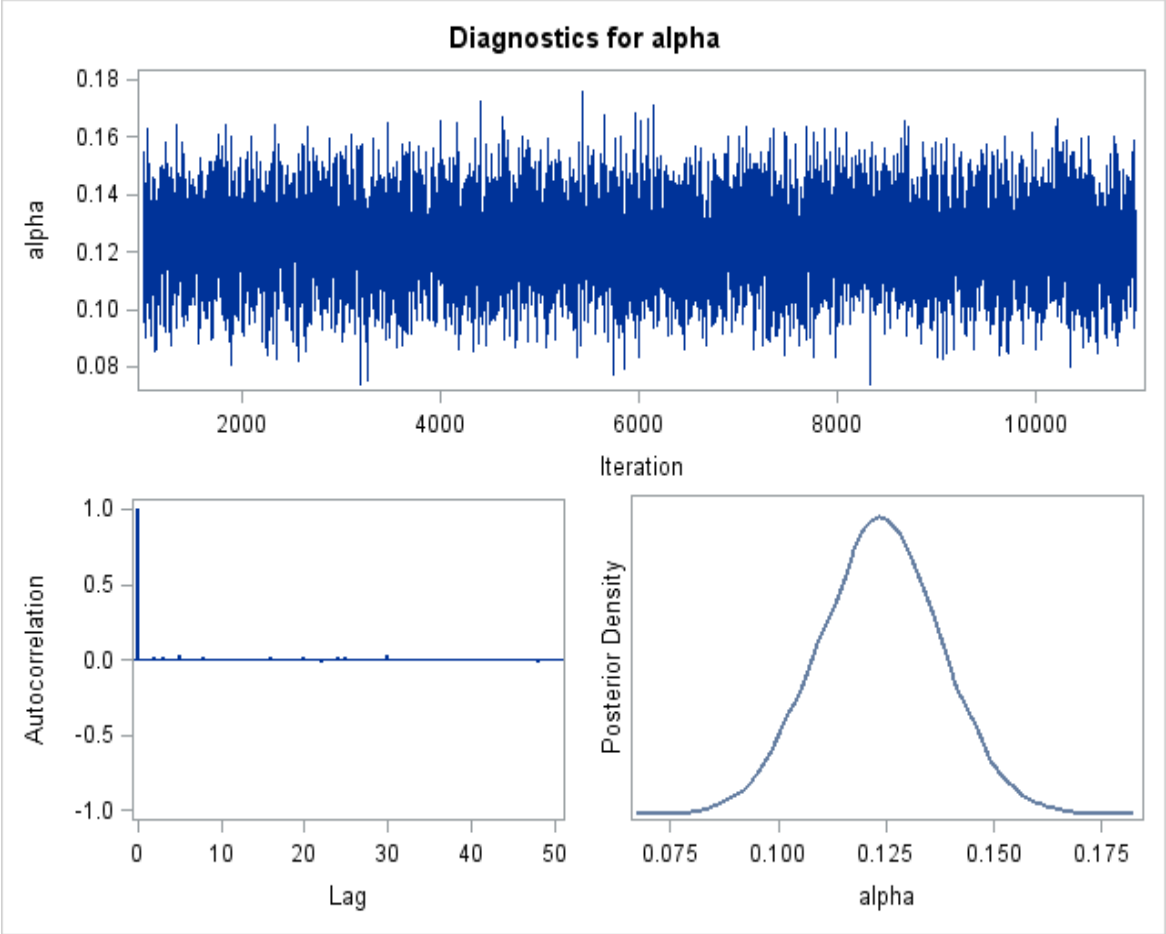
The MCMC Procedure



Priors – Pr_2

Intervention Type: Motives

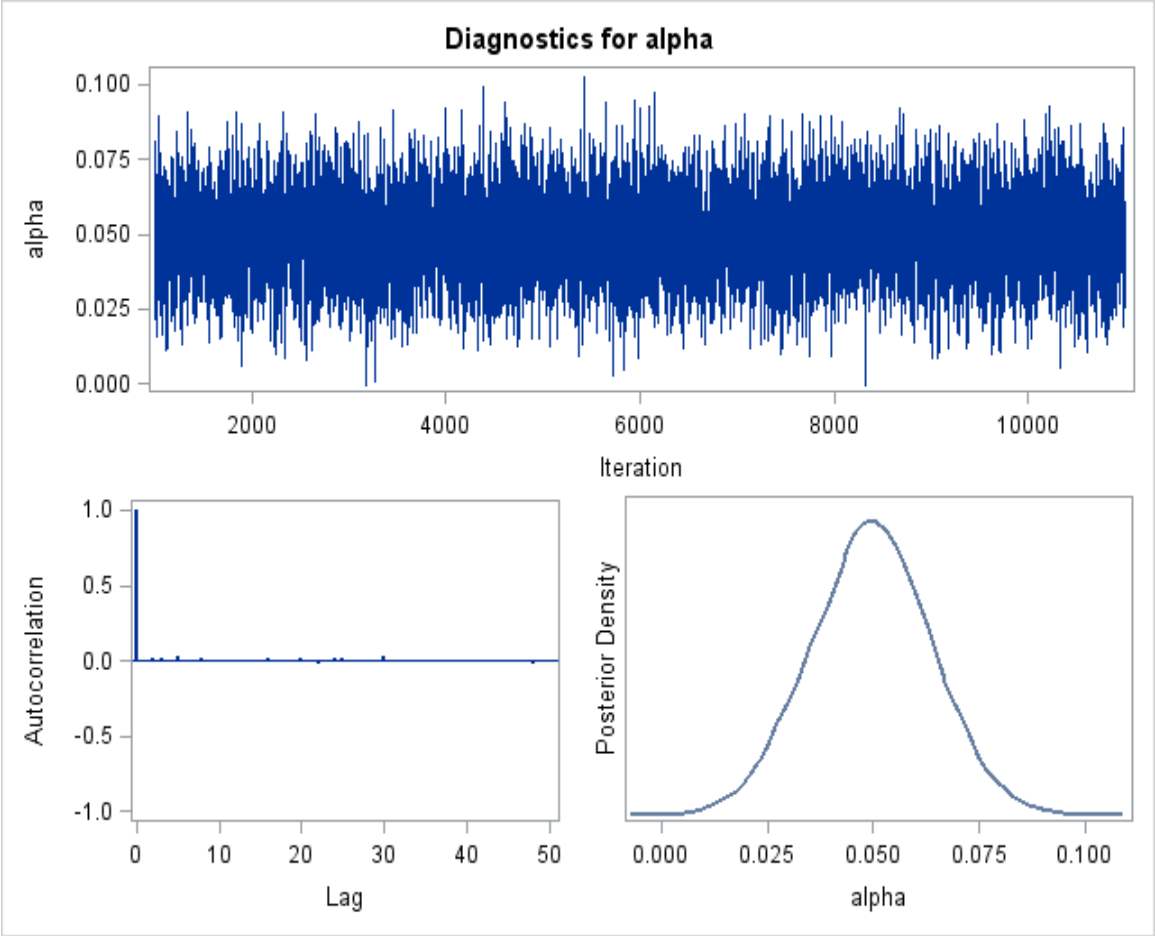
The MCMC Procedure



Posteriors – Pr_3

Intervention Type: Data

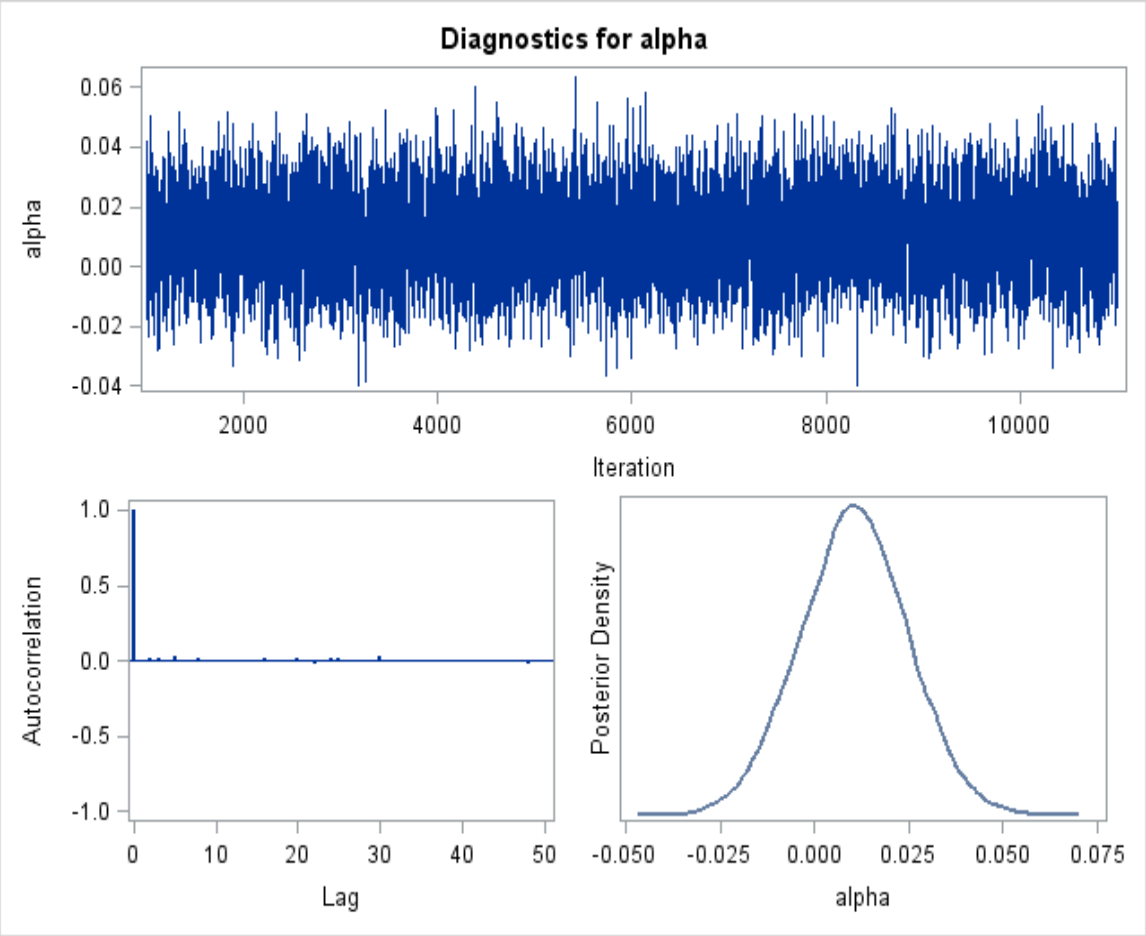
The MCMC Procedure



Posteriors – Pr_3

Intervention Type: Instrumentation

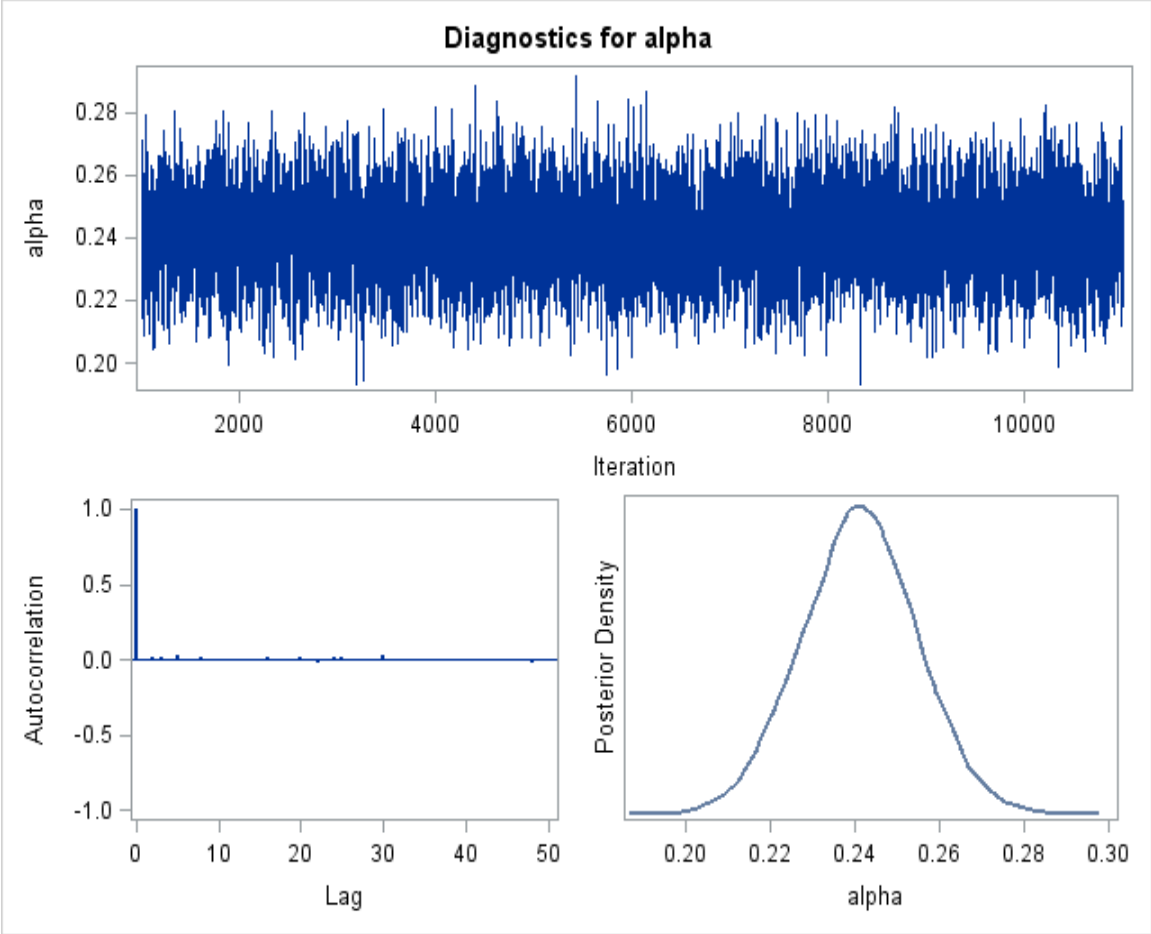
The MCMC Procedure



Posteriors – Pr_3

Intervention Type: Incentives

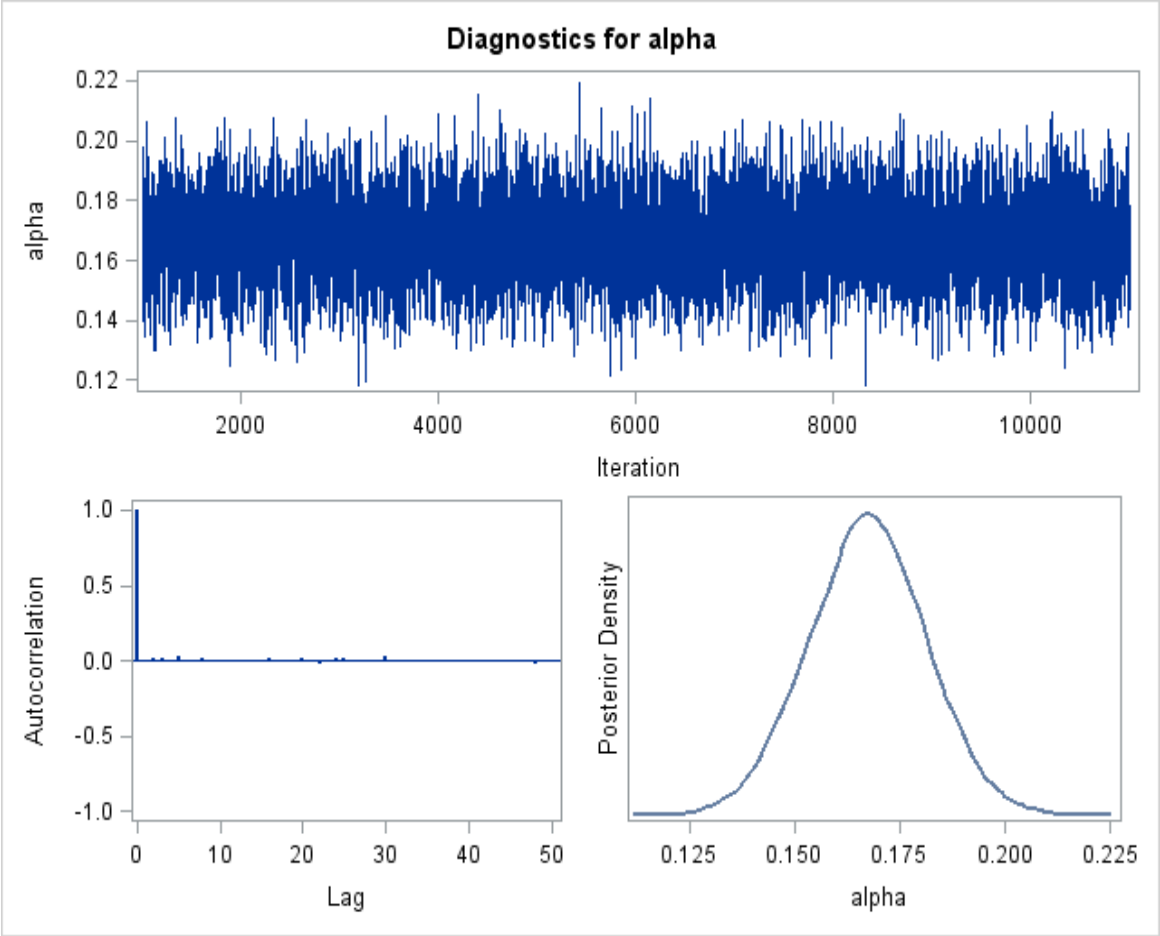
The MCMC Procedure



Posteriors – Pr_3

Intervention Type: Knowledge

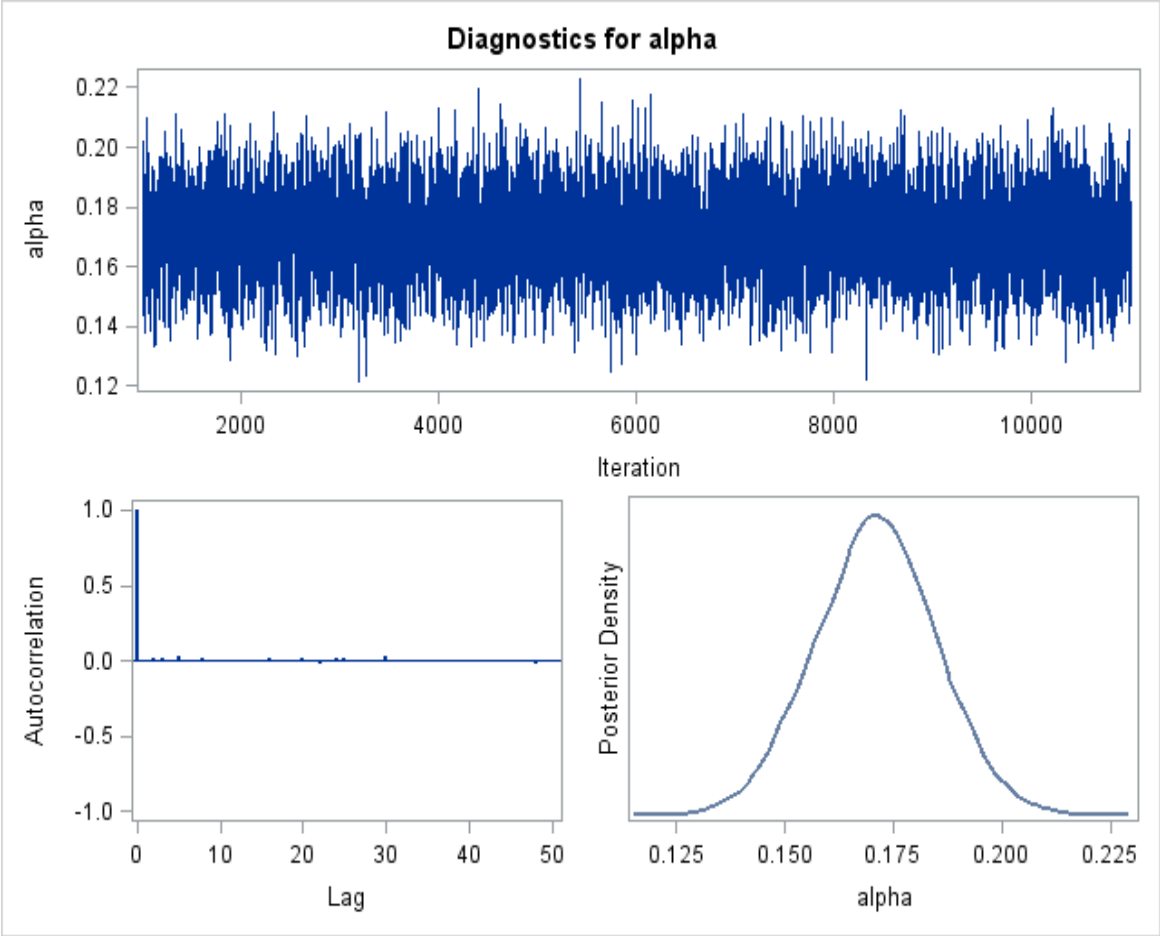
The MCMC Procedure



Posteriors – Pr_3

Intervention Type: Capacity

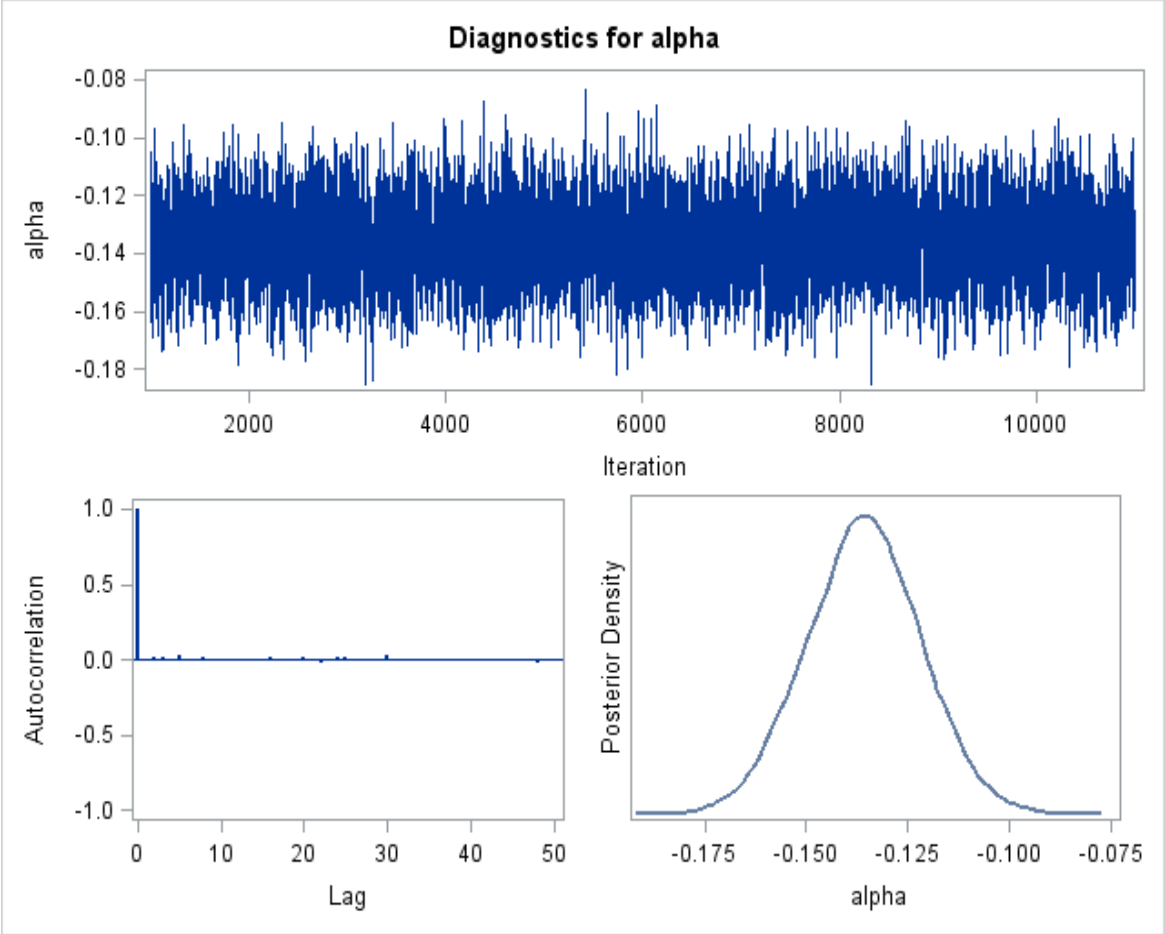
The MCMC Procedure



Posteriors – Pr_3

Intervention Type: Motives

The MCMC Procedure



REFERENCES

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-Based practice in psychology. *American Psychologist, 61*(4), 271–285.

Baker, G. E., Grubbs, A. B., & Ahern, T. M. (1990). Triangulation: Strengthening your best guess. *Performance Improvement Quarterly, 3*(3), 27-35.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44*(3), 211-233.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences, 30*(03), 241-254. doi:210.1017/S0140525X07001653.

Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, 58-80.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society, 53*, 370-418.

Beer, V. (1996). All of the intervention eggs were in one basket: A response to Swanson and Zuber. *Performance Improvement Quarterly, 9*(4), 79-81.

Berger, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association, 95*(452), 1269-1276.

Binder, C. (1995). Promoting HPT innovation: A return to our natural science roots. *Performance Improvement Quarterly, 8*(2), 95-113.

Blackburn, S. (2005) *Oxford Dictionary of Philosophy* (2nd ed.). Oxford, UK: Oxford University Press.

Bolstad, W. M. (2007). *Introduction to Bayesian Statistics* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Brethower, D. M. (1982). The total performance system. In R. M. O'Brien, A. M. Dickinson, & M. P. Rosow (Eds.), *Industrial behavior modification*, (pp. 350 – 369). New York: Pergamon Press.

Brethower, D. M. (2000). Integrating theory, research, and practice in human performance technology: Examples from behavioral, cognitive, and constructivist theories. *Performance Improvement, 39*(4), 33-43.

Brown, M. G. (1986). Determining the real causes of organizational performance problems. *Performance + Instruction, 25*(9), 13-15.

Bullock, D. (1973). Whatever happened to what's it's name, programmed instruction? Or a front-end/rear-end analysis of "fuzzies" as organizational goals. *NSPI Newsletter, 12*(4), 8-12.

Buxton, B. (1982). Training history: A practitioner's perspective. *Performance + Instruction, 21*(3), 20-24.

Buxton, B. (1984). Learning systems: Architecture and intervention. *Performance + Instruction, 23*(6), 8-12.

Chapman, C. R. (1973). Prior probability bias in information seeking and opinion revision. *The American Journal of Psychology, 86*(2), 269-282.

Chapman, D. D. (2004). Preferences of training performance measurement: A comparative study of training professionals and non-training managers. *Performance Improvement Quarterly, 17*(4), 31-49.

Christensen, T. K., & Osguthorpe, R. T. (2004). How do instructional design practitioners make instructional strategy decisions? *Performance Improvement Quarterly, 17*(3), 45-65.

Clark, R. E. (2003). Turning research and evaluation into results for ISPI. *Performance Improvement, 42*(2), 21-22.

Clark, R. E., & Estes, F. (1998). Technology or craft: What are we doing? *Educational Technology, 38*(5), 5-11.

Clark, R. E., & Estes, F. (2000). A proposal for the collaborative development of authentic performance technology. *Performance Improvement, 39*(4), 48-53.

Clark, R. E., & Estes, F. (2002). *Turning research into results: A guide to selecting the right performance solutions*. Atlanta, GA: CEP Press.

Conee, E. (2004). First things first. In E. Conee & R. Feldman (Eds.), *Evidentialism: Essays in epistemology* (pp. 11-36). Oxford: Clarendon Press.

Converse, D., & Weaver, C. L. (2008). Improving decision quality: A process for communicating data. *Performance Improvement, 47*(10), 19-24.

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.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*(1), 1-73.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: SAGE.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Darabi, A. (2003). Systems thinking and systematic methodology: A semi-empirical experience in support of the ISPI value proposition. *Performance Improvement, 42*(1), 17-23.

Davis, R. D., Latham, D. R., & Pitts, R. E. (1985). A program for teaching statistical process control to workers with limited mathematical skills: An intervention perspective. *Performance + Instruction, 24*(2), 22-24.

Dean, P. (1997). Editorial: Social systems science and the practice of performance improvement. *Performance Improvement Quarterly, 10*(3), 3-6.

Dean, P. J. (1995). Examining the practice of human performance technology. *Performance Improvement Quarterly, 8*(2), 68-94.

Delaplace, F., & Lescanne, P. (2009). HedN game, a relational framework for network based cooperation Retrieved March 20, 2010, from hal.archives-ouvertes.fr

Development, A. S. f. T. a. (2008). *State of the industry report* (pp. 32). Alexandria: ASTD.

Donaldson, S. I. (Producer). (2006, December 5, 2008). Thriving in the global Zeitgeist of accountability and evidence-based practice. [Streaming video of conference proceedings] Retrieved from http://web.cgu.edu/media/sbos/1%20-%20Intro%20-%20Zeitgeist%20of%20Evidence%20Based%20-%20Donaldon.mp4

Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2009). *What counts as credible evidence in applied research and evaluation practice*. Los Angeles, CA: Sage Publications, Inc.

Dufty, D. (2007). Bayes' Law in plain English. Retrieved from http://www.statgun.com /research/bayes-law.html

Earman, J. (1992). *Bayes or Bust?* Cambridge, MA: The MIT Press.

Efron, B. (1986). Why isn't everyone a Bayesian? *The American Statistician, 40*(1), 1-11.

Efron, B. (2005). *Modern science and the Bayesian-Frequentist controversy*. Stanford, CA: Stanford University.

Erikkson, L., & Hájek, A. (2007). What are degrees of belief? *Studia Logica, 86*, 183-213. doi: 10.1007/s11225-007-9059-4

Evans, J. S. B. T., Handley, S. J., Over, D. E., & Perham, N. (2002). Background beliefs in Bayesian inference. *Memory & Cognition, 30*(2), 179-190.

Farrington, J., & Clark, R. E. (2000). Snake oil, science, and performance products. *Performance Improvement, 39*(10), 5-10.

Feldman, R. (2003). *Epistemology*. Upper Saddle River: Prentice Hall.

Fitelson, B. (2001). *Studies in Bayesian confirmation theory.* (Unpublished doctoral dissertation). University of Wisconsin-Madison.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327-358.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378-382.

Foshay, R. (2000). Science and research applications in operational environments. *Performance Improvement, 39*(4), 44-47.

Gayeski, D. (1995). Changing roles and professional challenges for human performance technology. *Performance Improvement Quarterly, 8*(2), 6-16.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 513-525.

Gilbert, T. (1992). Foreword to the first edition. In J. Pershing (Eds.), *Handbook of Human Performance Technology,* (3rd ed.). pp. xxxi-xxxv. San Francisco: John Wiley & Sons.

Gilbert, T. F. (1978). *Human competence: Engineering worthy performance*. New York: McGraw-Hill Book Company.

Gilbert, T. F., & Gilbert, M. (1989). Performance engineering: Making human productivity a science. *Performance + Instruction, 28*(1), 3-9.

Goode, W. J. (1957). Community within a community: The professions. *American Sociological Review, 22*(2), 194-200.

Goodie, A. S. (1997). *Base-rate neglect under direct experience.* (Doctoral dissertation). Retrieved from ProQuest Information & Learning.

Goodie, A. S., & Fantino, E. (1996). Learning to commit or avoid the base-rate error.

Grant, D. A., & Moseley, J. L. (1999). Conducting a customer-focused performance analysis. *Performance Improvement, 38*(6), 15-18.

Guerra, I. J. (2001a). Performance improvement based on results: Is our field interested in adding value? *Performance Improvement, 40*(1), 6-10.

Guerra, I. J. (2001b). *A study to identify key competencies required of performance improvement professionals.* Florida State University, Tallahassee, FL.

Guerra, I. J. (2003). Asking and answering the right questions: Collecting useful and relevant data. *Performance Improvement, 42*(10), 24-28.

Guerra, I. J. (2003). Key competencies required of performance improvement professionals. *Performance Improvement Quarterly, 16*(1), 55-72.

Guerra-López, I. J. (2007). The impact evaluation process, part 1: Building a case for demonstrating the worth of performance improvement interventions. *Performance Improvement, 46*(7), 33-38.

Guerra-López, I. J., & Leigh, H. N. (2009). Are performance improvement professionals measurably improving performance? What PIJ and PIQ have to say about the current use of evaluation and measurement in the field of performance improvement. *Performance Improvement Quarterly, 22*(2), 97-110.

Hájek, A. (2007). Interpretations of probability. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* Palo Alto: Stanford University Press.

Hall, R. H. (1968). Professionalization and bureaucratization. *American Sociological Review, 33*(1), 92-104.

Hampton, J. M., Moore, P. G., & Thomas, H. (1973). Subjective probability and its measurement. *Journal of the Royal Statistical Society, 136*(1), 21-42.

Hannum, W. H. (2009). Training myths: False beliefs that limit the efficiency and effectiveness of training solutions, part 1. *Performance Improvement, 48*(2), 26-30.

Harless, J. H. (1995). Performance technology skills in business: Implications for preparation. *Performance Improvement Quarterly, 8*(4), 75-88.

Hawthorne, J. (2005). Degree-of-belief and degree-of-support: Why Bayesians need both notions. *Mind, 114*, 277-320.

Hempel, C. G. (1945a). Studies in the logic of confirmation (I.). *Mind, 54*(213), 1-26.

Hempel, C. G. (1945b). Studies in the logic of confirmation (II.). *Mind, 54*(214), 97-121.

Herem, M. A. (1979). Identifying causes of job performance deficiencies. *Improving Human Performance Quarterly, 8*(1), 53-62.

Hill, J., & Brethower, D. (1997). Ridding ourselves of "non-instruction." *Performance Improvement, 36*(8), 6-9.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association, 70*(350), 271-289.

Hoonakker, P., & Carayon, P. (2009). Questionnaire survey nonresponse: A comparison of postal mail and internet surveys. *International Journal of Human - Computer Interaction, 25*(5), 348-373.

Howson, C. (1995). Theories of probability. *The British Journal for the Philsophy of Science, 46*(1), 1-32.

Hripcsak, G., & Heitjan, D. F. (2002). Measuring agreement in medical informatics reliability studies. *Journal of Biomedical Informatics, 35*, 99-110.

Huglin, L., Johnson, L., & Marker, A. (2007). Research priorities in performance technology: A delphi study. *Performance Improvement Quarterly, 20*(1), 79-96.

Husted, Janice A., et al. "Methods for assessing responsiveness: a critical review and recommendations." Journal of clinical epidemiology 53.5 (2000): 459-468.

Hutchison, C. (1989). Moving from instructional technologist to performance technologist. *Performance + Instruction, 28*(9), 5-8.

Hutchison, C. S., Stein, F. S., & Carleton, J. R. (1996). Potential strategies and tactics for organizational performance improvement. *Performance + Instruction, 35*(3), 6-9.

Hutchison, C. S., & Stein, F. S. (1998). A whole new world of interventions: The performance technologist as integrating generalist. *Performance Improvement, 37*(5), 18-25.

Hybert, P. R. (2001). Give customers what they meant to ask for: Designing training systems at three levels. *Performance Improvement, 40*(9), 25-31.

Hybert, P. R. (2003). HPT, ISD: The challenge of clear boundaries in an evolving discipline. *Performance Improvement, 42*(2), 32-34.

Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation* (3rd ed.). San Diego, CA: EdITS.

ISPI. (2009). About ISPI Retrieved May 4, 2009, from http://www.ispi.org/content.aspx?id=208&linkidentifier=id&itemid=208

Jang, H. Y. (2008). *Themes and issues as reflected in human performance technology literature: A content analysis.* Indiana University.

Janis, I. L. (1972). Victims of Groupthink: A psychological study of foreign-policy decisions and fiascoes. Boston: Houghton-Mifflin.

Jeffrey, R. C. (1983). *The Logic of Decision* (2nd ed.). Chicago, IL: University of Chicago Press.

Jonassen, D. H. (1989). Performance analysis. *Performance + Instruction, 28*(4), 15-23.

Jones, W. P., & Kottler, J. A. (2006). *Understanding research: Becoming a competent and critical consumer*. Upper Saddle River, NJ: Pearson Education.

Joyce, J. (2003). Bayes' Theorem. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* Palo Alto, CA: Stanford University Press.

Kaplowitz, M. D., Hadlok, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly, 68*(1), 94-101.

Kastigar, G. (1991). Aligning and tracking a performance analysis project. *Performance + Instruction, 30*(4), 5-7.

Kaufman, R. (1985). Needs and motives. *Performance + Instruction, 24*(10), 1.

Kaufman, R. (2002). Commentary: Resolving the (often deserved) attacks on training. *Performance Improvement, 41*(6), 5-6.

Kaufman, R., & Clark, R. E. (1999). Re-establishing performance improvement as a legitimate area of inquiry, activity and contribution: Rules of the road. *Performance Improvement, 38*(9), 13-16.

Kaufman, R., Oakley-Browne, H., Watkins, R., & Leigh, D. (2003). *Strategic planning for success: Aligning people, performance, and payoffs*. San Francisco, CA: Pfeiffer.

Kaufman, R., & Valentine, G. (1989). Relating needs assessment and needs analysis. *Performance + Instruction, 28*(10), 10-14.

Kaufman, R., & Watkins, R. (2000). Getting serious about results and payoffs: We are what we say, do, and deliver. *Performance Improvement, 39*(4), 23-32.

Kaufman, R., Watkins, R., Sims, L., Crispo, N. S., Hall, J. C., & Sprague, D. E. (1997). Cost-consequence analysis: A case study. *Performance Improvement Quarterly, 10*(3), 7-21.

Kaufman, R., Watkins, R., Triner, D., & Smith, M. (1998). The changing corporate mind: Organizations, visions, mission, purposes, and indicators on the move toward societal payoff. *Performance Improvement Quarterly, 11*(3), 32-25.

Keeney, R. L. (1982). Decision analysis: An overview. *Operations Research, 30*(5), 802-838.

Kirschner, P. A., van Merrienboer, J., Sloep, P., & Carr, C. (2002). How expert designers design. *Performance Improvement Quarterly, 15*(4), 86-104.

Klein, J. D. (2002). Empirical research on performance improvement. *Performance Improvement Quarterly, 15*(1), 99-110.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes, 56*(1), 28-55.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*(01), 1-17.

Kolmogorov, A. (1956). *Foundations of the Theory of Probability* (2nd ed.). Retrieved from http://www.mathematik.com/Kolmogorov/index.html

Korth, S. J. (1997). Planning HRD interventions: What, why, and how. *Performance Improvement Quarterly, 10*(4), 51-71.

Korth, S. J. (2000). Creativity and the design process. *Performance Improvement Quarterly, 13*(1), 30-45.

Kramer, R. M. (1998). Revisiting the Bay of Pigs and Vietnam decisions 25 years later: How well has the groupthink hypothesis stood the test of time? *Organizational Behavior and Human Decision Processes*, *73*(2), 236-271.

Kukalis, S. (2009). Survey of recent developments in strategic management: Implications for practitioners. *International Journal of Management, 26*(1), 99-107.

Kusy, M. E., Jr. (1988). The effects of types of training on support of training among corporate managers. *Performance Improvement Quarterly, 1*(2), 23-30.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. doi: DOI:10.2307/2529310. JSTOR 2529310. PMID 843571

Langdon, D. (1992). Performance technology in three paradigms: Paradigm 2 - intervention identification. *Performance + Instruction, 31*(10), 1-5.

Langdon, D. (1997a). Are we doing what we say? *Performance Improvement, 36*(10), 5.

Langdon, D. (1997b). The performance technologist's role in interventions: An interview with Joe Harless. *Performance Improvement, 36*(9), 36-38.

Langdon, D. (1997). Selecting interventions. *Performance Improvement, 36*(10), 11-15.

Langdon, D. (2003). Should we conduct cause analysis or change-of-state analysis? *Performance Improvement, 42*(9), 8-13.

Langdon, D., & Whiteside, K. (1997). Introduction to the special issue on interventions. *Performance Improvement, 36*(10), 4.

Langdon, D. G., Whiteside, K. S., & McKenna, M. M. (1999). *Intervention resource guide: 50 Performance improvement tools*. San Francisco, CA: Pfeiffer.

Larson, M. B. (2005). Instructional design career environments: Survey of the alignment of preparation and practice. *TechTrends, 49*(6), 22-68.

Larson, M. B., & Lockee, B. B. (2004). Instructional design practice: Career environments, job roles, and a climate of change. *Performance Improvement Quarterly, 17*(1), 22-40.

Leigh, D. (2004). Editor's notes: The art of getting results. *Performance Improvement, 43*(6), 2-4.

Lewis, T. (2005). Towards artistry: A critique of the HRD performance paradigm and a suggested new model. *Performance Improvement Quarterly, 18*(2), 56-75.

Lindgreen, A., Swaen, V., & Johnston, W. J. (2009). Corporate social responsibility: An empirical investigation of U.S. organizations. *Journal of Business Ethics, 85*, 303-324.

Lindsey, J. D., & Cheek, E. H. (1986). Intervention strategies to reduce anxiety among students. *Performance + Instruction, 25*(10), 35-37.

Lovelady, L. (1984). Change strategies and the use of OD consultants to facilitate change. *Leadership and Organizational Development Journal, 5*(4), 2-12.

Lynch, S. M. (2007). *Introduction to applied bayesian statistics and estimation for social scientists*. New York, NY: Springer.

Mager, R. F. (1992). The "T" in "PT" has got to go. *Performance + Instruction, 31*(2), 57.

Mager, R. F., & Pipe, P. (1997). *Analyzing performance problems or you really oughta wanna* (3rd ed.). Atlanta: The Center for Effective Performance.

Mann, E. (1996). *A case study of instructional design practices: Implications for designers.* Paper presented at the National Convention of the Association of Educational Communications and Technology, Indianapolis, IN.

Marker, A., Huglin, L., & Johnson, L. (2006). Empirical research on performance improvement: An update. *Performance Improvement Quarterly, 19*(4), 7-22.

Marrelli, A. (2010). Data collection. In R. Watkins & D. Leigh (Eds.), *Handbook of Improving Performance in the Workplace: Selecting and Implementing Performance Interventions* (Vol. 2, pp. 792-822). San Francisco: Pfeiffer.

Mattson, B. W. (2003). The effects of alternative reports of human resource development results on managerial support. *Human Resource Development Quarterly, 14*(2), 127-151.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730-740.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*(1/2), 103-113.

Medsker, K., & Fry, J. (1992). Toward a performance technology curriculum. *Performance + Instruction, 31*(2), 53-56.

Medsker, K., Hunter, P., Stepich, D., Rowland, G., & Basnet, K. (1995). HPT in academic curricula: Survey results. *Performance Improvement Quarterly, 8*(4), 6-21.

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of "unstructured" decision processes. *Administrative Science Quarterly, 21*(2), 246-275.

Mitchell, D. (1993). Just plain talk: On getting some respect for our craft. *Performance + Instruction, 32*(9), 40-41.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640-647.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology, 66*, 337-344.

Muir, M., Watkins, R., Kaufman, R., & Leigh, D. (1998). Cost-consequences analysis: A primer. *Performance Improvement, 37*(4), 8-17, 48.

Munley, M. (2003). From a training request to performance consulting. *Performance Improvement, 42*(7), 18-25.

Munn, T. (2005). Commentary: Can science help? *Performance Improvement, 44*(2), 5-8.

Namboodiri, N. K. (1972). Experimental designs in which each subject is used repeatedly. *Psychological Bulletin, 77*(1), 54-64.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. [Review]. *Review of General Psychology, 2*(2), 175-220.

Nickols, F. (2005). Performance as intervention (performer as interventionist). *Performance Improvement, 44*(1), 9-12.

Nutt, P. C. (1984). Types of Organization Decision Processes. *Administrative Science Quarterly, 29*(3), 414-450.

O'Driscoll, T. (2003). Learning from history: Chronicling the emergence of human performance technology. *Performance Improvement, 42*(6), 9-22.

Ormerod, R. J. (1997). The design of organizational intervention: Choosing the approach. *Omega, International Journal of Management Science, 25*(4), 415-435.

Ouwersloot, H., Nijkamp, P., & Rietveld, P. (1998). Errors in probability updating behaviour : Measurement and impact analysis. *19*(5), 535-563.

Parmigiani, G., & Inoue, L. (2009). *Decision theory: Principles and approaches*. Chichester, UK: John Wiley and Sons, Ltd.

Patten, M. (2001). *Questionnaire research: A practical guide* (2nd ed.). Glendale, CA: Pyrczak Publishing.

Pershing, J. (Ed.). (2006). *Handbook of human performance technology: Principles, practices, potential* (3rd ed.). San Francisco, CA: Pfeiffer.

Pershing, J. A., Cheng, J., & Foong, K. P. (2006). International Society for Performance Improvement professional practices survey: A report. *Performance Improvement, 45*(7), 39-47.

Pershing, J. A., Ji-Eun Lee, & Jing Li Cheng. (2008a). Current status, future trends, and issues in human performance technology, part 1: Influential domains, current status, and recognition of HPT. *Performance Improvement, 47*(1), 9-17.

Pershing, J. A., Ji-Eun Lee, & Jing Li Cheng. (2008b). Current status, future trends, and issues in human performance technology, part 2: Models, influential disciplines, and research and development. Performance Improvement, 47(2), 7-15.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 785-794.

Quinn, R. E., & McGrath, M. R. (1982). Moving beyond the single-solution perspective: The competing values approach as a diagnostic tool. *The Journal of Applied Behavioral Science, 18*(4), 463-472.

Ramsey, F. P. (1926). Truth and probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays* (pp. 156-198). New York: Harcourt, Brace and Company. 1999 electronic edition retrieved on August 31, 2009.

Rassin, E. (2008). Individual differences in susceptibility to confirmation bias. *Netherlands Journal of Psychology, 64*(2), 87-93.

Raven, B. H. (1998). Groupthink, bay of pigs, and watergate reconsidered.*Organizational Behavior and Human Decision Processes*, *73*(2), 352-361.

Reiser, R. A. (2001). A history of instructional design and technology: Part II: A history of instructional design. *ETR&D, 49*(2), 57-67.

Reiser, R. A., & Dempsey, J. V. (Eds.). (2007). *Trends and Issues in Instructional Design and Technology* (2nd ed.). Upper Saddle River, NJ: Pearson Education Inc.

Richey, R., Fields, D., Foxon, M., Roberts, R. C., Spannaus, T., & Spector, J. M. (2001). Instructional design competencies: The standards

Richey, R. C. (1986). Theoretical bases of instructional design: Learning and instruction *The theoretical and conceptual bases of instructional design* (pp. 57-73). London: Kogan Page.

Richmond, S. (1984). The interaction of art and science. *Leonardo*, 81-86.

Robinson, D. G., & Robinson, J. C. (1996). *Performance Consulting: Moving Beyond Training*. San Francisco, CA: Berrett-Koehler Publishers.

Robinson, D. G., & Robinson, J. C. (2006). Commentary. Performance consulting: The art and the science. *Performance Improvement, 45*(4), 5-8.

Romme, A. G. L., & Damen, C. M. (2007). Toward science-based design in organizational development: Codifying the process. *Journal of Applied Behavioral Science, 43*(1), 108-121.

Rosenberg, M. J., Coscarelli, W. C., & Smith Hutchinson, C. (1999). The origins and evolution of the field. In H. D. Stolovitch & E. J. Keeps (Eds.), *Handbook of Human Performance Technology: Improving Individual and Organizational Performance Worldwide* (2nd ed.) (pp. 24-46). San Francisco: Jossey-Bass/Pfeiffer.

Rossett, A., & Tobias, C. (1999). A study of the journey from training to performance. *Performance Improvement Quarterly, 12*(3), 31-43.

Rowland, G. (1992). What do instructional designers actually do? An initial investigation of expert practice. *Performance Improvement Quarterly, 5*(2), 65-86.

Ruckdeschel, C., Yarter, M., Riveccio, M. A., Cortes, I., & Cookson, M. S. (1998). Beyond instructional systems: A performance technology degree. *Performance Improvement, 37*(3), 22-26.

Rummler, G. A. (1982). Instructional technology and organization performance: An historical perspective. *Performance + Instruction, 21*(3), 9-11.

Rummler, G. A. (2004). *Serious performance consulting: According to Rummler*. Silver Spring: International Society for Performance Improvement.

Rummler, G. A., & Brache, A. B. (1995). *Improving performance: How to manage the white space on the organization chart* (2nd ed.). San Francisco: Jossey-Bass.

Runyon, R. P., Coleman, K. A., & Pittenger, D. J. (2000). *Fundamentals of Behavioral Statistics* (9th ed.). Boston: McGraw-Hill.

Salmon, W. C. (1967). *The foundations of scientific inference*. Pittsburgh, PA: University of Pittsburgh Press.

Sanchez, J. C. (1993). The long and thorny way to an organizational taxonomy. *Organization Studies, 14*(1), 73-92.

Schwen, T. M., Leitzman, D. F., Misanchuk, E. R., & Foshay, W. R. (1979). Instructional development: A social intervention? *Improving Human Performance Quarterly, 8*(4), 242-258.

Schwitzgebel, E. (2008). Belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Palo Alto, CA: Stanford University Press.

Shoemaker, H. A. (1972). The president reports: What does NSPI stand for? *NSPI Newsletter, 11*(9), 8-9.

Shrock, S., & Coscarelli, W. C. (1981). Some sources of the art, craft, and science views of instructional development. *Performance + Instruction, 20*(7), 26-48.

Shuttleworth, M. (2009). Counterbalanced measures design Retrieved from http://www.experiment-resources.com/counterbalanced-measures-design.html

Sincoff, M. Z., & Owen, C. L. (2004). Content guidelines for an undergraduate human resources curriculum: Recommendations from human resources professionals. . *Journal of Education for Business, 80*(2), 80-85. doi: Document ID: 786056431

Skyrms, B. (1986). *Choice and Chance*. Belmont, CA: Wadsworth, Inc.

Sleesman, D. J., Conlon, D. E., McNamara, G., & Miles, J. E. (2012). Cleaning up the big muddy: A meta-analytic review of the determinants of escalation of commitment. *Academy of Management Journal*, *55*(3), 541-562.

Sleezer, C. M. (1992). Needs assessment: Perspectives from the literature. *Performance Improvement Quarterly, 5*(2), 34-46.

Staw, B. M. (1976). Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational behavior and human performance*, *16*(1), 27-44.

Staw, B. M. (1981). The escalation of commitment to a course of action. *Academy of Management Review*, *6*(4), 577-587.

Stigler, S. M. (1975). Studies in the history of probability and statistics. XXXIV Napoleonic statistics: The work of Laplace. *Biometrika, 62*(2), 503-517.

Stigler, S. M. (1982). Thomas Bayes's Bayesian inference. *Journal of the Royal Statistical Society. Series A (General), 145*(2), 250-258.

Stolovitch, H. (2007). The development and evolution of human performance improvement. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Stolovitch, H., & Keeps, E. J. (2006). Foreword to the third edition. In J. Pershing (Ed.), *Handbook of Human Performance Technology* (pp. xiii-xix). San Francisco, CA: Pfeiffer.

Stolovitch, H., & Keeps, E. J. (2008). Selecting solutions to improve workplace performance. In E. Biech (Ed.), *ASTD Handbook for Workplace Learning Professionals* (pp. 147-169). Alexandria, VA: ASTD Press.

Stolovitch, H. D. (2000). Human performance technology: Research and theory to practice. *Performance Improvement, 39*(4), 7-16.

Sugrue, B., & Stolovitch, H. (2000). Report of 1999 ISPI Symposium—Appropriate inquiry in human performance technology. *Performance Improvement, 39*(1), 33-36.

Svenson, R. (2006). Requirements: The bridge between analysis and design. In J. Pershing (Ed.), *Handbook of Human Performance Technology* (3rd ed.) (pp. 223-237). San Francisco, CA: Pfeiffer.

Swanson, R. A., & Zuber, J. A. (1996). A case study of a failed organization development intervention rooted in the employee survey process. *Performance Improvement Quarterly, 9*(2), 42-56.

Terwee, C. B., Dekker, F. W., Wiersinga, W. M., Prummel, M. F., & Bossuyt, P. M. M. (2003). On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research*,*12*(4), 349-362.

Tessmer, M. (1993). *Planning and conducting formative evaluations*. London: Kogan Page Limited.

Tessmer, M., & Wedman, J. (1992,). *What designers do, don't do, and why they don't do it*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Thomas, M. N. (2006). Evidence-practice partnership. *Performance Improvement, 45*(6), 8-12.

Thomas, M. N. (2007, May). *Evidence essentials: Tips for evidence-based practices.* Paper presented at the 2007 International Performance Improvement Conference: Performance Beyond Borders, San Francisco, CA.

Tosti, D. T. (2005). The "big five:" The evolution of the performance systems model. *Performance Improvement, 44*(9), 9-13.

Tosti, D. T., & Kaufman, R. (2007). Who is the "real" father of HPT? *Performance Improvement, 46*(7), 5-8.

Tovar, M., Gagnon, F., & Schmid, R. (1997). Development of a consultation profile perceived as successful by human performance technology consultants. *Performance Improvement Quarterly, 10*(3), 67-83.

Trussell, N. (2008) *Contingent Incentives.* In Lavrakas, P. J. (Ed.). *Encyclopedia of Survey Research Methods. Sage Publications*. Thousand Oaks, USA: Sage Publications.

Tsai, M. H., & Young, M. J. (2010). Anger, fear, and escalation of commitment. *Cognition and Emotion*, *24*(6), 962-973.

Tschan, F., Semmer, N. K., Gurtner, A., Bizarri, L., Spychiger, M., Breuer, M., & Marsch, S. (2009). Explicit reasoning, confirmation bias, and illusory transactive memory: A simulation study of group medical decision making. *Small Group Research, 40*(3), 271-299.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

US Department of Commerce (2012). *2012 NAICS Definitions.* Retrieved from http://www.census.gov/cgi-bin/sssd/naics/naicsrch?chart=2012

Van Tiem, D. M. (2004). Interventions (solutions) usage and expertise in performance technology practice: An empirical investigation. *Performance Improvement Quarterly, 17*(3), 23-44.

Van Tiem, D. M., Moseley, J. L., & Dessinger, J. C. (2000). *Fundamentals of Performance Technology: A Guide to Improving People, Process, and Performance*. Silver Spring, MD: ISPI.

Van Tiem, D. M., Moseley, J. L., & Dessinger, J. C. (2001). *Performance improvement interventions: Enhancing people, processes, and organizations through performance technology*. Silver Spring, MD: ISPI.

Visscher-Voerman, I., & Gustafson, K. L. (2004). Paradigms in theory and practice of education and training design. *ETR&D, 52*(2), 69-89.

Wagenmakers, E. J., Lee, M., Lodewyckx, T., & Iverson, G. J. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses in psychology* (pp. 181-207). New York: Springer.

Wallsten, T. S., & Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science, 29*(2), 151-173.

Watkins, R. (2007a). Designing for performance, Part 1: Aligning your HPT decisions from top to bottom. *Performance Improvement, 46*(1), 7-13.

Watkins, R. (2007b). Designing for performance, Part 2: Selecting your performance technologies. *Performance Improvement, 46*(2), 9-15.

Watkins, R., & Kaufman, R. (1996). An update on relating needs assessment and needs analysis. *Performance + Instruction, 35*(10), 10-13.

Watkins, R., & Leigh, D. (Eds.). (2010). *Handbook of improving performance in the workplace: Selecting and implementing performance interventions* (Vol. 2). San Francisco: Pfeiffer.

Watkins, R., Leigh, D., & Kaufman, R. (2000). A scientific dialogue: A performance accomplishment code of professional conduct. *Performance Improvement, 39*(4), 17-22.

Watkins, R., Leigh, D., Platt, W., & Kaufman, R. (1998). Needs assessment: A digest, review, and comparison of needs assessment literature. *Performance Improvement, 37*(7), 40-53.

Watkins, R., & Wedman, J. (2003). A process for aligning performance improvement resources and strategies. *Performance Improvement, 42*(7), 9-17.

Wayne, L. (2009, May 29). A promise to be ethical in an era of immorality, *New York Times*. Retrieved from http://www.nytimes.com/2009/05/30/business /30oath.html?_r=1&scp=2&sq=harvard%20business%20MBA%20code&st=cse

Wedman, J., & Tessmer, M. (1993a). Instructional designers' decisions and priorities: A survey of design practice. *Performance Improvement Quarterly, 6*(2), 43-57.

Wedman, J., & Tessmer, M. (1993b). Instructional designers' decisions and priorities: A survey of design practice. *Performance Improvement Quarterly, 6*(2), 43-57.

Wedman, J. F., & Graham, S. W. (1998). Introducing the concept of performance support using the performance pyramid. *Journal of Continuing Higher Education, 46*(3), 8–20.

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*(679-700).

Wellins, R. S., & Rothwell, W. J. (2008). Competencies: The foundation of our profession. In E. Biech (Ed.), *ASTD Handbook for Workplace Learning Professionals* (pp. 807-822). Alexandria, VA: ASTD Press.

Westgaard, O. (1997). Describing a performance improvement specialist: The heurist. *Performance Improvement, 36*(6), 10-15.

Whatever happened to what's it's name, programmed instruction? Or a front-end/rear-end analysis of "fuzzies" as organizational goals. (1973). *NSPI Newsletter, 12*(4), 8-12.

Winer, L. R., Vasquez-Abad, J. A., & Tessmer, M. (1994). *Enriching the layers of necessity model*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Witte, E. (1972). Field research on complex decision-making processes—The phase theorem. *International Studies of Management and Organization*, 156-182.

Witteman, C., & Renooij, S. (2003). Evaluation of a verbal–numerical probability scale. *International Journal of Approximate Reasoning, 33*, 117–131.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology, 59*(5), 555-561.

ABSTRACT

THE ROAD TAKEN:
A BETWEEN AND WITHIN SUBJECTS STUDY OF INTERVENTION SELECTION DECISIONS BY PERFORMANCE IMPROVEMENT PROFESSIONALS

by

HILLARY N. LEIGH

December 2014

Advisor: Dr. Ingrid Guerra-López

Major: Instructional Technology

Degree: Doctor of Philosophy

Background: In performance improvement, intervention selection is a complex decision that ought to be based on the best available evidence. Despite this, there is little research into what sources of evidence are used in intervention selection and what changes in belief occur during performance improvement professionals' decisions. Framed in decision theory, this study aims to resolve these problems. **Methods:** Sixty-one Certified Performance Technologists completed a dynamic, Web-delivered questionnaire where they provided a general assessment of intervention success (Pr_1), then responded to 12 performance improvement scenarios; by selecting an intervention, providing a prior probability, receiving additional evidence, giving a posterior probability (Pr_3), indicating whether the initial intervention was still preferred and making a subsequent choice if not. **Results:** Repeated measures ANOVA showed significant interaction between time and evidential agreement for probability assessments (p <.001). No effects were shown for scientific nature of evidence. Informed Bayesian analyses

showed only main effects (evidential agreement and time). Factorial MANOVA found significant effects for evidential agreement on Pr_3 and changes between prior and posterior probability (*l)* (p <.001). Marginally mixed effects were noted for scientific nature of evidence. Normal (*Z)* approximation revealed subjects tended to stick with their initial intervention choice (p<.0001) and only scientific evidence was associated with this action (r_s=-0.3160, p<.0001); informed Bayesian analyses revealed contrary findings. Binary logistic regression illustrated Pr_3 (*OR*=1.085*)* and *l* (*OR*=3.792*)* are good models for changes of mind (p<.0001, Max-rescaled R^2=.75). When subjects did change their minds, no differences in self-reported familiarity on initial interventions existed (p = 0.085), but familiarity was significantly lower for subsequently preferred interventions (p = 0.003). Post hoc paired t-tests showed higher levels of familiarity with selected interventions than their non-selected counterparts (p <.0001). No significant correlations occurred between familiarity and Pr_3, four analyses yielded correlations for general and prior assessments of likely interventions success. Informed Bayesian analyses illustrated dramatically different results, specifically, 15 of the 18 correlational analyses between self-reported familiarity and assessments of likely intervention success were significant. Repeated measures ANOVA showed no significant effects of practice on the repeated probability measures (p = 0.806) and post hoc ANOVA showed that randomized blocks were similar (p <.0001) and no differences between them (p =.201, p =.604, p =.072).

Discussion: These findings bolster the long-standing concern about the technical nature of performance improvement and practitioners are strongly encouraged to approach intervention selection as a *decision*, where their intervention preferences, beliefs of likely

success are carefully adjudicated on the basis of the evidence they obtain. Future research with other types of performance improvement practitioners, replication studies, longitudinal, structural equation modeling, externally verifiable probabilities, and natural environments are recommended.

AUTOBIOGRAPHICAL STATEMENT

Hillary Leigh has been working to improve organizational performance for 10 years. Her interest in Performance Improvement grew out of a desire to live in a better world and an appreciation for organizations' unique potential (and obligation) to make it so. Before pursuing a Doctoral degree, she obtained a Master's degree and certificate in Performance Improvement from University of Michigan-Dearborn and a Bachelor's degree from Oakland University. While completing her graduate studies, she consulted in healthcare, educational, retail, and non-profit organizations to select, develop, implement and evaluate a variety of performance interventions. During her studies, she authored or co-authored 2 book chapters, 2 journal articles, 1 book review; presented or co-presented at 8 conference presentations, including an invited presentation on changing minds at ISPI's 2012 Research-to-Practice symposium; supported research evaluating the use of technology in schools for two funded grants; and taught graduate-level courses in Instructional Design, Performance Consulting, and Evaluation.

Her current professional work as an internal consultant for Kaiser Permanente's Southern California Permanente Medical Group involves the management of a large-scale patient care experience survey. This includes the development of an integrated web-enabled questionnaire and the administration of an incentive program related to key measures included in the instrument. She is rewarded by the commitment of the organization to "doing the right thing", the complexity of the work, and the dedication of her colleagues. Happy to call Pasadena, California home, she looks forward to returning to guilt-free enjoyment of her hobbies of wine-tasting, cooking, reading fiction, and cycling—in that order.