5-6-2018

How Experts Judge Creativity: A Field Study of the Assessment of Creative Output

Michael Robert Seyle

PERMISSION TO BORROW

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, copy from, or publish this dissertation may be granted by the author or, in her absence, the professor under whose direction it was written or, in his absence, by the Dean of the Robinson College of Business. Such quoting, copying, or publishing must be solely for scholarly purposes and must not involve potential financial gain. It is understood that any copying from or publication of this dissertation that involves potential gain will not be allowed without written permission of the author.

Michael Robert Seyle

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:
Michael Robert Seyle
915 Lorena Street
Mamaroneck, New York 10543

The director of this dissertation is:

Pam Scholder Ellen
J. Mack Robinson College of Business
Georgia State University
Atlanta, GA 30302-4015

How Experts Judge Creativity: A Field Study of the Assessment of Creative Output

by

Michael Robert Seyle

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Executive Doctorate in Business

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY

ROBINSON COLLEGE OF BUSINESS

2018

ACCEPTANCE

This dissertation was prepared under the direction of the *MICHAEL ROBERT SEYLE* Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Dr. Pam Scholder Ellen (Chair)

Dr. Steve D. Olson

Dr. Edward E. Rigdon

ACKNOWLEDGEMENTS

I am deeply indebted to and sincerely appreciate the hard work and support of my Dissertation Committee Chair, Dr. Pam Scholder Ellen, and the members of my Dissertation Committee, Dr. Steve D. Olson and Dr. Edward E. Rigdon. Their guidance, direction and mentorship were invaluable to the completion of the research and the dissertation. All three are also wonderful people with a passion for research, teaching and living a full life. I am very lucky to have them as instructors, mentors and professional colleagues during and after this academic journey.

I would also like to thank my professors, fellow doctoral students, and the staff of the Executive Education program at the Robinson College Business of Georgia State University. I have never met a more dedicated, caring and professional group of individuals, and I am honored to have experienced such an amazing program with them. In particular, I want to thank Dr. Lars Matthiessen, Academic Director, and Maury Kalnitz, former Director of the Executive Doctorate Program, for their confidence and guidance. Their vision and leadership helped create an outstanding opportunity for executives to develop new skills and understanding, and to contribute to both academia and practice in a meaningful way.

Last, but by no means least, I want to thank my wonderful wife, Michelle, and my children, Kathryn, Emily, Haley, Jack and Gia, for all of their support, love and understanding as I pursued my dream of another advanced degree. I hope my curiosity and love of learning inspires them to reach for and achieve their dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

How Experts Judge Creativity: A Field Study of the Assessment of Creative Output

by

Michael Robert Seyle

August 2018

Chair: Pam Scholder Ellen

Major Academic Unit: Executive Doctorate in Business

Creativity is a fundamental component of innovation and critical for long-term business success. Identifying the products and ideas that are most creative, and therefore worthy of further development and investment, is an essential part of the creative process. However, experimental research into creativity over the past 20 years has yielded inconsistent and contradictory results. Moreover, this same research has shown that organizations struggle to identify their most creative products and ideas for further development. Critics suggest organizational creativity research may suffer from measurement misspecification due to a misalignment between existing construct definitions of creativity as a response that is both "novel *and* useful" and experimental studies that use only a single item, *creativity,* to measure creative output. This research investigates whether the theorized misalignment may be due, in part, to the research use of judges with little or no experience in the creative domain and their failure to understand the criteria, approaches, and techniques that expert judges actually use to assess creative output. To better understand how these issues may affect research results, this research utilized Naturalistic Decision Making field-study methods to investigate how expert judges assess the creative output of experienced professionals in the setting of a creativity awards contest. Through a series of interviews, observations, think-aloud protocols, and simulations

with expert judges of creativity award contests, this research identifies six factors experts use to assess professional level creative output, and uncovers the processes, approach, and challenges involved in the real-world assessment of creative products and ideas. Recommendations for how assessing creativity can be improved in research and practice are discussed as well as suggestions for future research.

INDEX WORDS: Creativity, Intuition, Analysis, Experts, Creative Output, Measurement, Field Study, Contests, Judgment, Decision Making

I INTRODUCTION

I.1 Background

The importance of creativity to business success and the global economy continues to garner attention in both business press and academia. Expanding on Drucker's concept of the "knowledge-based" organization, Florida argues human creativity has become "the decisive source of competitive advantage" for businesses and "the ultimate economic resource" for organizations (Florida, 2002). In 2008, the Conference Board reported that corporate executives view creativity as a key capability organizations need to succeed:

> U.S. employers rate creativity and innovation among the top five skills that will increase in importance over the next five years and rank it among the top challenges facing CEOs (Conf. Board, 2008).

A 2010 survey of 1,500 U.S. corporate CEOs extended the significance of creativity to the C-suite, stating, "Creativity is now the most important leadership quality for success in business, outweighing even integrity and global thinking" (IBM, 2010).

Creativity is viewed as an essential element and precursor to innovation that drives organizational success and economic prosperity (Amabile & Khaire, 2008). The importance of creativity to both business and global economies is reflected in the impact creative industries have on domestic production. Creative industries, such as the arts and cultural goods and services, are estimated to contribute more than $500 billion to the U.S. Gross Domestic Product, or more than 3.2% of current-dollar GDP, compared to other industries, like travel and tourism, that generate only 2.6% of GDP (U.S. Bureau of Economic Analysis, 2013). Advertising creativity alone is estimated to generate almost $200 billion in U.S. gross domestic output, or 20% of all arts and cultural goods and services (US BEA 2013). Likewise, in the United

Kingdom, creative industries supply more than 1.5 million jobs, which amounted to more than 5% of the U.K.'s total employment in 2010 (UK Dept. Culture Media and Sport, 2011).

I.2 Problem Statement

Given creativity's importance to business success and global economic output, it is unsurprising that individual and organizational creativity has been the subject of increasingly intensive research in the fields of management, psychology, education, and the arts over the past 60 years (Runco, 2004). Despite this intense academic interest across disciplines, however, understanding of creativity remains fragmented into specialized subfields, often resulting in inconsistent, and conflicting, research findings (Amabile, 2012). Recently, scholars have raised concerns that these inconsistent and contradictory creativity research results may be due to construct measurement misspecification, incomplete methodological approaches, or poor construct validity (Sullivan & Ford, 2010; Montag et al., 2012). Many of these concerns relate to how creativity is defined, both as a construct and operationally, as well as the processes utilized in assessing creative output in research studies (Montag et al., 2012).

Most extant creativity research has utilized consensual judgment methods to assess the "overall creativity" of products or ideas as a unitary construct or single item. This approach contradicts the numerous existing construct definitions of creativity as either a two-item or a multi-faceted construct (Sullivan & Ford, 2010). The majority of published research to date also has relied upon experimental studies that employ children, students, and individuals with little or no experience or training in creativity as study participants (Montag et al., 2012). These inexperienced experimental subjects raise concerns about the quality of the items assessed and the validity of their results (Byron & Khazanchi, 2012). Moreover, the majority of studies have utilized educators, students, and other individuals with little or no established creativity expertise

as judges to measure the creative output of study subjects. In short, it appears existing creativity research may suffer from the triple-threat combination of low construct validity, measurement misspecification, and incomplete assessment methods. To investigate the nature and extent of these potential shortcomings, and gain insight into the processes, criteria, and approaches used to assess creative products and ideas in practice, this research examined how established domain experts make judgments of the creative output of professionals in a real world setting.

I.3 Statement of Purpose and Research Questions

The purpose of this field study was to investigate the processes and criteria expert judges use in assessing the output creativity of those who regularly create products and ideas in creative industries that are subsequently submitted to award contests. Understanding how domain experts make judgments about the level of professionals' creativity in situations where judges face significant time, uncertainty, and ambiguity constraints provides useful insights into how creativity is assessed in real-world situations. Such insights may also: (1) increase the field's understanding of the challenges creativity researchers face in assessing creative output in experimental studies, (2) identify the key criteria real world judges use in assessing creativity; (3) suggest more effective assessment methods to improve the overall reliability and validity of creativity research, and (4) provide ideas that organizations can use to better discern the most promising creative product and ideas for development and investment.

To explore these issues, this research examines how domain experts judge the creativity of entries in a professional award program by investigating the following research questions:

1. What criteria do expert judges of creativity awards contests report using to assess the creativity of products and ideas; in particular, do experts use the "novel and useful" definition of creativity that is employed in most creativity research or some other criteria?

2. What processes do experts use as they judge creativity in contest settings; are the processes involved in creativity contests similar to or different from those used in the experts' professional workplace settings?

3. What types of cognitive processing are involved in expert judgment of creative products and ideas; do experts use intuition, rational analysis, a combination of both, or something else when judging creativity contests?

4. What do experts say are the differences between how experts and non-experts make judgments of creativity; what kinds of mistakes might a novice make in judging creativity?

5. What biases and heuristics do expert judges acknowledge encountering when judging creativity in real-world situations, and how do they attempt to deal with them?

I.4 Research Approach

After receiving approval for the study from the university's institutional review board, this research began investigating the steps, processes, and experiences of experts judging entries in a creativity awards contest. Identified expert participants were all professionals in various creative industries with significant experience in assessing creative output in organizational settings as well as having acted as judges in numerous regional, national, and international creativity award programs.

The main method of data collection involved in-depth interviews of these experts. Following an engaged scholarship research approach, the investigation originated with a pilot study that conducted open-ended interviews of managers in different creative industries. This approach revealed the nature of creativity assessment in organizational settings and identified key challenges facing organizations that seek to exploit their most creative products and ideas. To gain further understanding into the processes involved in judging creative products and ideas,

the researcher subsequently spent three days immersed in a creativity awards event, observing expert judges during the judging process, collecting information about the awards program, and interviewing contest judges and the organizers of the awards program. In-person interviews of participants were conducted using Applied Cognitive Task Analysis techniques in an effort to uncover non-conscious routines and steps in the judging process. Several participant judges also took part in simulated judging exercises and think-aloud protocols during judging of contest entries to provide a better explanation of their cognitive effort. After the awards event, additional phone interviews were conducted with domain experts who had served as judges for the same awards program in prior years. A total of 12 in-depth interviews of expert judges were administered during and after the three-day judging event, and five judges also participated in simulation and think-aloud exercises during the judging event. In addition to the interviews, simulations and observations, archival data on the judges' numerical scores across ten creativity award contests over five years were obtained from the contest organizers. This data was obtained in non-attributable format for use in testing the judges' inter-rater reliability and comparing rater agreement within and across awards contests.

The combination of interviews, many involving Applied Cognitive Task Analysis with simulations and think-aloud exercises, along with direct observation of the judging event, and a longitudinal analysis of contest scores, did not allow for complete triangulation of data, but did help provide a deep analysis of creativity assessment in a field setting. A comprehensive review of the literature on the assessment of creativity and expert judgment, the pilot study, and direct observations helped shape the collection of additional data from past judges. Preliminary coding classifications were developed in advance of interviews, simulations and think-aloud exercises, and coding was refined as the research progressed, taking into account data previously collected

and the frameworks underlying the data analysis. To help insure reliability of the coding process, two independent researchers check-coded several interview transcripts and disagreements were discussed until consensus was reached on a coding scheme that was used by the researcher for the remaining transcripts.

I.5 Researcher Background and Assumptions

The researcher who conducted this study is employed as an executive at a global architecture and design organization and has more than a decade of direct experience working with professionals engaged in generating creative products and ideas on a regular basis. This background and practical experience provides a unique understanding about creativity in the workplace and the challenges organizations face in the development of creative and innovative output. This background, however, also tends to color and potentially limit the researcher's perceptions and beliefs about creative work and how creativity is judged. As a result, the researcher took additional steps to identify and address potential preconceptions and biases that might affect the neutrality of the research, including peer reviews of the research design and process, consultation with academics and consultants not connected to the research, journaling about the decisions and choices made in the study's design and analysis, and informing each of the research participants of the researcher's position, such that they could raise any concerns they might have during the research process. To help strengthen the validity of the research and address the subjectivity of the researcher, interview candidates were not prescreened as to their viewpoints and were selected by the contest organizers based solely on their qualifications as domain expert judges of numerous creativity contests. Multiple data sources were sought to help identify diverse perspectives and possible alternative explanations for the findings and information discovered. Information that potentially contradicted any researcher assumptions or

existing theoretical concepts was preserved and considered separately, to prevent forcing the data to fit any preconceptions.

Based on past experience and understanding of the challenges businesses face in assessing creativity, as well as concerns raised in the literature about judgments of creative output, the following three assumptions grounded this research. First, creativity is a social construct that is dependent upon the experiences, values, and utility ascribed to it by individuals who hold a position of authority, respect, or recognition within a particular domain or industry. As such, creativity cannot be measured objectively through scientific analysis or accurately described as a physical property—whether something is considered creative is an intersubjective assessment or shared meaning developed through interaction with others and by means of comparison, expectation and personal aesthetic judgment. Second, although all individuals can be creative in some way and at some period, those who are employed in creative industries to generate creative output have special skills, training, talent, motivation, supports, and experience that are necessary to inform creative products and ideas for business. This second assumption leads to a third, wherein those who have been employed as creative professionals for many years and who have been identified by their peers as masters of their craft, such that they are selected to judge the creativity of other professionals, have achieved a level of experience to be considered "experts" in their domain.

I.6 Rationale and Significance

The rationale for this study arises from the challenges business leaders and managers of creative industries face in identifying the most creative products and ideas for further development and investment. In addition, academic research into creativity has struggled to produce consistent and reliable findings about the nature of creativity, its antecedents and

consequences, and methods to motivate individuals and increase creative output. One possible reason for inconsistent, and often contradictory, results may stem from how the dependent variable of "creative outcomes" is measured and assessed. A deeper understanding of how experts judge creativity in field settings should lead to improved processes for assessing creative output and lend new insights into how to measure, increase, and expand creativity.

I.7 Summary of Remaining Chapters

The remaining chapters of this dissertation cover the following areas:

- **Chapter 2 - Literature Review:** This section provides a general review of the literature on creativity and the assessment of creative outcomes. It examines previous research on the definitions and processes of creativity, measurement of creative output in experimental studies, and the decision-making aspects of creative evaluation and assessment. This chapter highlights concerns about possible measurement misspecification and the validity and reliability of existing quantitative research. The review also generally explores the research on expert decision-making processes, particularly expert intuition and the heuristics and biases that accompany its use. The review establishes that a lack of research has explored how creativity is evaluated in real-world settings and the processes domain experts use in judging creativity; in doing, it identifies the gap in literature this study is designed to address. This section also describes the two main theories of information processing relating to judgment and decision making, Cognitive-Experiential Self-Theory (CEST) and Cognitive Continuum Theory (CCT), explores how each theory applies to previous research, and the constructs and predictions of each theory about how decisions are made and outcomes are evaluated. This section illustrates how CEST and CCT differentially view the role and impact of intuition and analysis in decision-making, the nature and impact of heuristics and human cognitive biases, and the reliability of judgments under different conditions. This chapter also provides an overview of Naturalistic Decision Making (NDM) as a framework for investigating and understanding judgment and decision making of experts in field settings.

- **Chapter 3 - Research Methodology:** This chapter justifies the use of a qualitative, process focused case study approach, as it seeks to answer how experts judge creativity where the researcher has no control over activities, behaviors, and events of interest. This chapter also discusses the engaged scholarship approach to research as a means of increasing the relevance of the study by including information and perspectives from key stakeholders. The process used to select domain experts with significant experience in

judging and evaluating creativity in professional settings is also discussed. This chapter also provides an outline of the data collection, reduction, and triangulation strategies utilized to both increase understanding of the context and improve validity by incorporating: (1) multiple sources of evidence, (2) a case study database, and (3) a chain of evidence. It subsequently explains the methods used to analyze the various types of data obtained.

- **Chapter 4 – Findings**

  Chapter 4 presents the key findings of the in-depth interviews, simulations and think-aloud exercises, field observations of an actual creativity awards judging event, and the results of inter-rater reliability testing of judges' numerical contest scores. Extensive illustrative quotes from participant interviews are provided to allow the reader an opportunity to explore participants' views in their own words.

- **Chapter 5 – Discussion**

  This section provides an analysis of the findings and a synthesis of the data in an effort to make sense of the meaning of the data. It additionally explores how the data relates to the theoretical frameworks applied in this research study. Possible linkages and logical inferences suggested by the data are explored, and the main contributions of the study to theory and practice are discussed.

- **Chapter 6 – Conclusions and Recommendations**

  This final chapter provides the researcher's conclusions based on study findings and analysis, as well as recommendations for changes to research methods and practice based on the results of the study. Recommendations for future research are also discussed.

II LITERATURE REVIEW

II.1 Introduction

The purpose of this field study was to investigate the processes and criteria expert judges use in assessing the creativity of output by individuals who regularly create products in creative industries. Specifically, the research sought to understand how domain experts make judgments about the level of creativity of products and ideas, in situations where judges face extreme time pressure, uncertainty, and ambiguity. Before conducting the study, critical reviews of the literature on creativity research, construct measurement, decision making and expert cognition were conducted. The literature review informed the research design, data collection, and analysis phases of this study.

This review of creativity literature traces the approaches, methods, results, and challenges of research into the nature and elements of creativity; in particular, the ways in which creative outcomes are measured and assessed. Moreover, as discernment of creative outcomes involves many aspects of judgment and decision making, this review explores the constructs and theories relating to decision making, particularly as it concerns experts, rational choice, and intuition. Lastly, literature relating to how individuals make decisions in field settings is reviewed.

II.2 Review of Literature

Academic study of creativity began to flourish after J. P. Guilford's 1950 presidential address to the American Psychological Association (Mumford, 2003; Amabile, 2012). However, for the first 30 years of inquiry, creativity was considered solely a psychological "trait" or quality, consisting of unique attributes and "divergent" thinking abilities only possessed by gifted individuals (Amabile, 2012). As a result, almost all creativity research until the mid-

1990s focused on the inherent differences and abilities of exceptionally creative individuals (Mumford, 2003).

Over the past 20 years, creativity research has undergone a number of significant shifts in both the development of theory and the use of various research methodologies (Mumford, 2003). This research has also expanded from psychology to sociology, education, economics, and management (Runco, 2004). As a result of its multi-disciplinary focus, the scope of understanding creativity's processes, antecedents, and consequences has grown exponentially, but not without having paid a price (Hennessey & Amabile, 2010). The lack of inter-disciplinary research has resulted in fragmented, often conflicting and confusing, findings and the development of numerous competing theories of creativity, how it functions, how to measure it, and how to improve creative thinking and outcomes (Hennessey & Amabile, 2010).

II.2.1 *Definitions of Creativity*

What creativity is, and what it is not, hangs as the mythical albatross around the neck of scientific research on creativity (Prentky, 2000: 97).

One of the most significant challenges to creativity research the lack of a consistent definition. More than 50 years ago, Rhodes reported having identified 40 different definitions of creativity (Rhodes, 1961). Twenty-seven years later, a review of published psychological research uncovered more than 60 different definitions for the concept (Taylor, 1988). These various definitions reflect the growing, but fragmented, academic interest in the study of creativity from different domains and research perspectives.

When used as an adjective, the word "creative" is defined in the Oxford Dictionary as, "relating to or involving the use of the imagination or original ideas to create something" or "having good imagination or original ideas." As a noun, creative is defined as, "a person whose

job involves creative work." As a verb, to "create" is defined as meaning to "bring (something) into existence," or "cause (something) to happen as a result of one's actions." Creativity also has been defined as "inventiveness; the use of imagination or original ideas to create something."[1] Synonyms of creativity include cleverness, imagination, ingenuity, inventiveness, originality, resourcefulness, inspiration, and vision (Roget, 2013).

In creativity research, the term "creative" has developed numerous inherent meanings over the past 65 years that can refer to a person, group or organization, or to the process, products, and environment in which people work (Batey, 2012). For example, when creativity is viewed as an aspect of a *person*, the definition might arise from the traits and abilities inherent in creative genius (Runco, 2004; Runco & Jaeger, 2012). If the focus is on *process*, a definition might relate the mental activities and procedural steps involved in a divergent thinking exercise (Mumford, 2003).

The lack of a shared definition is certainly not due to a shortage of effort or attention. In 1996, Amabile attempted to define creativity from a social-psychological approach stating, "a product or response will be judged as creative to the extent that (1) it is both a novel and an appropriate, useful, correct, or valuable response to the task at hand, and (2) the task is heuristic rather than algorithmic" (Amabile, 1996: 35). Sternberg and Lubart (1999) defined creativity more objectively as "the ability to produce work that is both novel (i.e., original, unexpected) and appropriate (i.e., useful, adaptive concerning task constraints)" (Sternberg & Lubart, 1999:

---

[1] The origin of the word "create" is believed to be of late Middle English in the sense of "form out of nothing" and relating to "a divine or supernatural being" originally from the Latin *creat-* "produced," from the verb *creare*. "However, the verb form derives from a Proto-Indo-European root *kerh2*, which is believed originally to have meant, 'grow' and in Latin its original meaning was 'to make (something) grow.'" The past participle of *creare* is "to make, bring forth, produce, beget" which is related to *crescere*, or "arise, grow." (http://www.etymonline.com/index.php?term=create, *accessed February 9, 2015*).

3). This seeming confluence of conceptual understanding lead Mumford to conclude: "Over the course of the last decade, however, we seem to have reached a general agreement that creativity involves the production of novel, useful products" (Mumford, 2003: 110). However, despite these various attempts to capture the essence of creativity, many authors still believe a widely accepted or inter-disciplinary definition of creativity remains elusive and that the lack of a workable definition hampers efforts to formally identify and measure the construct (Sullivan & Ford, 2010; Batey, 2012; Montag et al., 2012; Simonton, 2012).

This study investigated the assessment of creative individuals' output (Montag, et al., 2012), as opposed to creative processes, individual traits, or the environments surrounding creative activity,; accordingly,the research necessarily focused only on creative products and ideas. Thus, for the purposes of the present research, and for the reasons outlined in the next section, this study adopts as a starting point Amabile's proffered definition of creative output as a product or response that is both novel and appropriate to the task, where the task is heuristic rather than algorithmic (Amabile, 1996; 2012).

II.2.2 *Creativity Research*

To understand the divergent views of creativity and why a standard definition of creativity is necessary, it is important to trace the development of creativity research over the past 30 years. According to the componential theory of creativity, developed by Amabile in 1983 and refined in the years since, creativity occurs in the combination of three "within-individual" components and one external component. The three internal components are: (1) domain-relevant skills, i.e., expertise or knowledge in the relevant domain or domains; (2) creativity-relevant process abilities, i.e., cognitive abilities and personality attributes that enable novel thinking; and (3) task motivation, specifically intrinsic motivation to engage in the task.

The single external component of the theory is the surrounding social or workplace environment the individual acts within (Amabile, 1983; 2012).

Amabile's componential theory therefore predicts, "creativity should be at its highest when an intrinsically motivated person with high domain expertise and high skill in creative thinking works in an environment high in supports for creativity" (Amabile, 2012: 3). Domain-relevant skills relate to the knowledge, experience, skills, and abilities of the particular technical or professional arena where an individual has gained expertise. This domain-relevant expertise provides the background the individual will likely draw upon to develop new combinations and to evaluate the viability of the various options created (Amabile, 1983). Creativity-relevant process abilities include "a cognitive style and personality characteristics that are conducive to independence, risk-taking, and taking new perspectives on problems, as well as a disciplined work style and skills in generating ideas" (Amabile, 2012: 3).

Task, or intrinsic, motivation has received the most scholarly attention in recent years, and the inconsistent, and often contradictory, research findings in this area frame many of the current definitional dilemmas. The intrinsic principle of motivation posits "people are most creative when they feel motivated primarily by the interest, enjoyment, satisfaction, and challenge of the work itself" (Amabile 2012: 4). However, the external factors that explicitly motivate the organization or the individual—so-called extrinsic motivators—as well as other workplace, mood, and compensation factors, can have either a positive or negative effect on an individual's intrinsic motivation and creativity (Oldham & Cummings, 1996; George & Zhou, 2007).

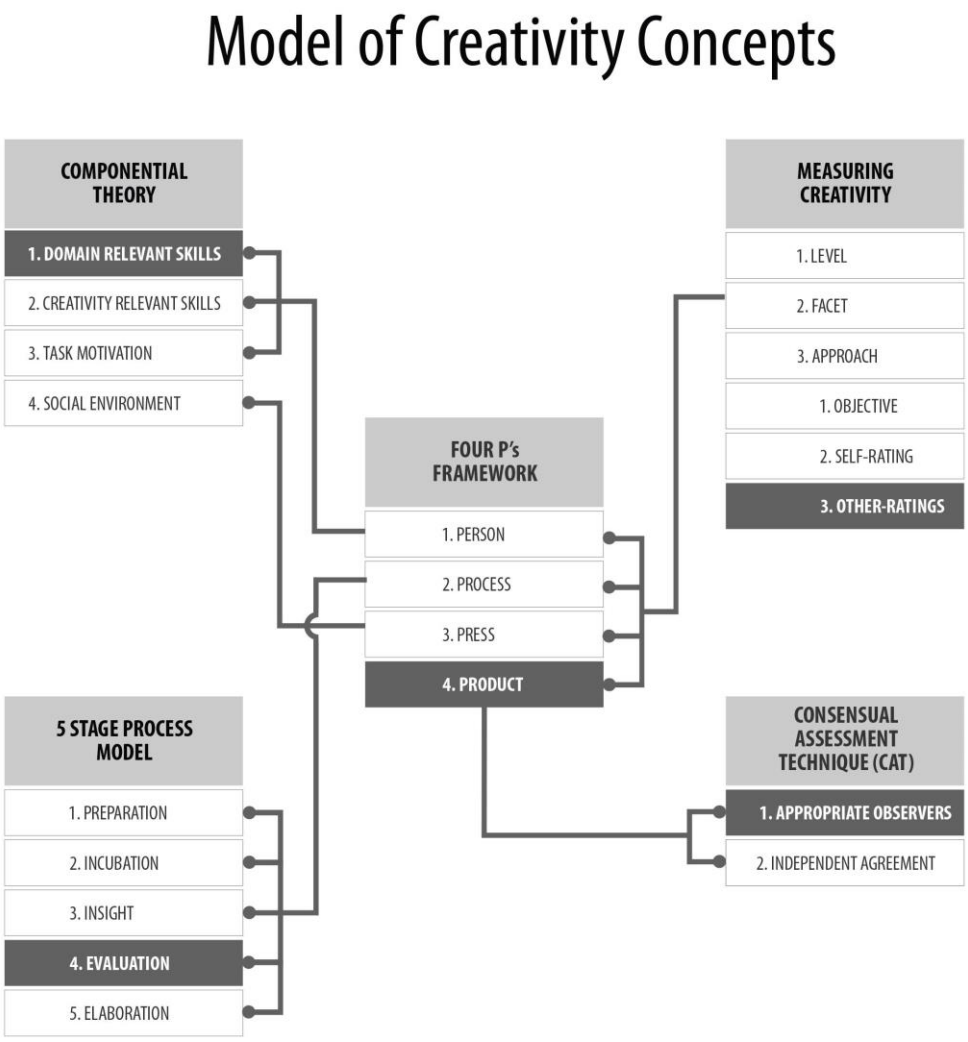
One of the most widely used frameworks for understanding the concept of creativity is the "Four P's" model, a schematic representation that divides creativity into four categories of

person, *process*, *place,* and *product* (Runco, 2004). Initially developed by Rhodes (1961), the Four P's identify each aspect of creativity as an attribute of: (1) the *Person*—i.e., some individuals tend to be more creative than others; (2) the *Process*—i.e., particular types of thinking, perception and behaviors are likely to produce creativity more readily than others; (3) *Place*—i.e., physical and social environmental factors can have the effect of increasing or decreasing creativity in individuals; and, (4) *Product*—i.e., some ideas, solutions and designs can be judged as more creative than others (Rhodes, 1961; Runco, 2004).

One of the most comprehensive models of the *Process* aspect of creativity is that developed by Csikszentmihalyi (1996), and consists of five stages: (i) preparation, (ii) incubation, (iii) insight, (iv) evaluation, and (v) elaboration. *Preparation* involves the preparatory work that focuses an individual's mind on the problem and explores the problem's dimensions. *Incubation* is the stage where the problem is internalized into the unconscious mind and nothing appears externally to be happening. *Insight* arises when the creative person begins to make connections between previously unassociated concepts and a creative idea arises from non-conscious processing into conscious awareness. *Evaluation* occurs when the newly developed idea is mentally challenged and tested for its appropriateness to the problem or goal. Finally, *elaboration* is the stage where the idea is further refined and then applied.

The *Product* category of the Four P's creativity model and the *evaluation* stage in Csikszentmihalyi's (1996) five-stage model provide the focal points for this research study. The *Product* dimension focuses on the creative outcomes and results of the creative process, specifically whether they are judged as more or less creative. The evaluation stage is crucial in creativity assessment as it provides the first opportunity to objectively assess the fitness of

generated ideas during the creative process. For organizations, the evaluation stage serves as a critical juncture where products and ideas are assessed for potential value and further development. An error in assessment at this stage can prevent creative ideas and products from being realized or result in unnecessary development effort and excessive cost. Businesses must also be able to assess the creativity of fully developed ideas and products before incorporating them into to strategic plans or introducing them to the marketplace. In research, the assessment of creative products or outcomes at either the



Adapted from Amabile, 1983, 2012; Runco, 2004; Csikzentmihalyi, 1996; and Batey, 2012.
Shaded items highlight focus areas of present research.

evaluation stage or conclusion of the creative process has been called “the bedrock of all studies of creativity” (Cropley & Cropley, 2008; MacKinnon 1978: 187).

Figure 1 Schematic Model of Creativity Concepts

II.2.3 *Creative Outcomes*

Similar to the challenges of defining creativity overall, the definitions of creative products and outcomes also have varied considerably over the past 50 years. Brogden and Sprecher (1964) defined a *product* simply as a physical object, a theoretical system, an equation, or a technique. Research studies using this definition attempt to explain the differences between creative and ordinary products (Santanen, Briggs, & Vreede, 2004). Jackson and Messick (1965) defined a *creative* product as being “unusual” when compared to other products: it is appropriate to the context of the situation; it shifts the constraints and boundaries of the situation; and it condenses both simplicity and complexity in such a way that the product may at first appear simple but is in reality quite complex. Condensing these definitions, Amabile et al. (1986) operationally defined a product as creative if it was both novel and appropriate in response to a non-algorithmic task (Santanen et al., 2004).

However, the past 30 years of research using these definitions of creativity and creative products has resulted in conflicting, and often contradictory, experimental outcomes across numerous disciplines (Batey, 2012; Hennessey & Amabile, 2010). For example, in the social-psychological study of creativity, particularly with regard to its motivational dimension, researchers have struggled to find common ground in methods to increase creativity (Hennessey & Amabile, 2010). A significant body of research shows extrinsic motivators, such as financial rewards or recognition, have a detrimental effect on creative output by increasing a sense of being controlled and reducing intrinsic motivation, a fundamental aspect of creativity according

to the componential theory (Amabile, 1996; Hennessey, 2003). Researchers following a behavioral perspective, however, have developed substantial empirical evidence showing how extrinsic rewards and recognition can increase creativity, particularly in the workplace, by focusing individuals on the need for creative ideas, without any resulting detrimental impact on intrinsic motivation (Cameron & Pierce, 1994; Eisenberger & Shanock, 2003; Eisenberger & Aselage, 2009).

Similarly, research into the effects one's work environment has on creativity has failed to reach consensus. For example, the effect of time pressure on creativity has been shown to have a detrimental impact on creativity in some studies (Amabile et al., 1996), while other studies have found that individuals respond positively to time pressures and produce greater levels of creativity (Madjar & Oldham, 2002). Studies also have shown a complicated relationship between time pressure, the person, and the situation, resulting in an inverted U-shape relationship that is mediated by personality (Baer & Oldham, 2006). Likewise, summaries of research into the related concepts of feedback, monitoring by supervisors, and evaluation of work product have shown these environmental factors to have quite different effects on creative output depending on how information is presented and processed by the individual (Zhou, 2003). In organizational research, some studies have shown the use of clear overall goals can increase motivation and creative output (Amabile, Hennessey & Grossman, 1996), while many others suggest such external expectations have a direct negative impact on creativity (Shalley & Perry Smith, 2001; West, 2002).

A number of meta-analyses and reviews of extant research also have failed to coalesce the disparate findings surrounding various approaches to creativity research, stating that "few solid conclusions regarding creativity can be drawn" (see Byron & Khazanchi, 2012; Oldham &

Baer, 2006; Shalley, Zhou, & Oldham, 2004; Montag, et al., 2012). For example, Byron & Khazanchi (2012) reviewed more than 60 experimental and field studies applying either Learned Industriousness Theory (LIT) or Self-Determination Theory (SDT) and concluded that neither theory completely accounted for the inconsistent and contradictory results of the various studies reviewed (Byron & Khazanchi, 2012). The authors also identified a number of study limitations that may have contributed to a lack of validity and reliability, including that the majority of the research subjects were children, students, or individuals who were not required or expected to perform creatively on a routine basis. Additionally, most of the studies did not occur in creative or professional settings, and the few studies situated in workplace settings did not focus specifically on creative employees (Byron & Khazanchi, 2012). As a result, the conclusions reached by LIT and SDT research cannot extend to questions of workplace creativity among professionals or employees in creative occupations. The authors also point out most of the tasks involved in the 60 studies reviewed offered little opportunity for high levels of creativity, explaining, "the studies in this analysis were more likely to employ tasks that were relatively low in complexity" and concluding the "proposed models may better explain incremental creativity than radical creativity" (Byron & Khazanchi, 2012: 825). In conclusion, the authors urge further research into the nature of creativity in the workplace using experienced creative employees as participants.

Another recent review of the creativity literature offered a different perspective to explain the inconsistent results of existing research. Montag, et al. (2012) criticized most creativity research approaches as unnecessarily focused and organized according to the antecedents of creativity, which may have resulted in incorrect conclusions and inconsistencies among prior studies and the inability to develop generalizations from those results (Montag et al, 2012).

Montag and colleagues instead propose a new framework for organizing and understanding workplace creativity constructs based on workplace performance literature. Instead of thinking of creativity as a unitary construct, they contend creativity should be considered a research domain with multiple constructs. The authors hypothesize at least two separate sets of constructs may exist within the creativity domain, i.e., Creative Performance Behaviors (CPB) and Creative Outcome Effectiveness (COE), with differential definitions and causal directions between them. The competing perspectives, inconsistent results, and contradictory findings of creativity studies over the past few decades have caused some commentators to question the appropriateness of the various frameworks, theories, and methodologies used in studies of creativity, as well as whether the definitions utilized are appropriate and complete, and whether the construct of creativity itself needs to be reconsidered (Sullivan & Ford, 2010; Montag et al., 2012).

II.3 Measurement of Creative Output

As outlined above, how creativity is measured depends in large part on the construct and operational definitions used as well as the research perspective employed. Separating creative behaviors from creative output as proposed by Montag et al. (2012) requires an understanding of the methods that have been used to measure creative output. Batey (2012) developed a heuristic, multi-level framework to account for different levels of analysis, facets, and approaches to measuring creativity. Within this taxonomic framework, shown in the diagram from Batey's article reproduced below, creativity can be analyzed from an individual, team, organization, or culture level by considering the four facets of creativity—i.e., traits, process, press, or product. Measures of creativity under this heuristic model can employ: (1) objective measures, (2) subjective assessments through self-reports, or (3) external appraisals or ratings by subject matter experts (Batey, 2012).

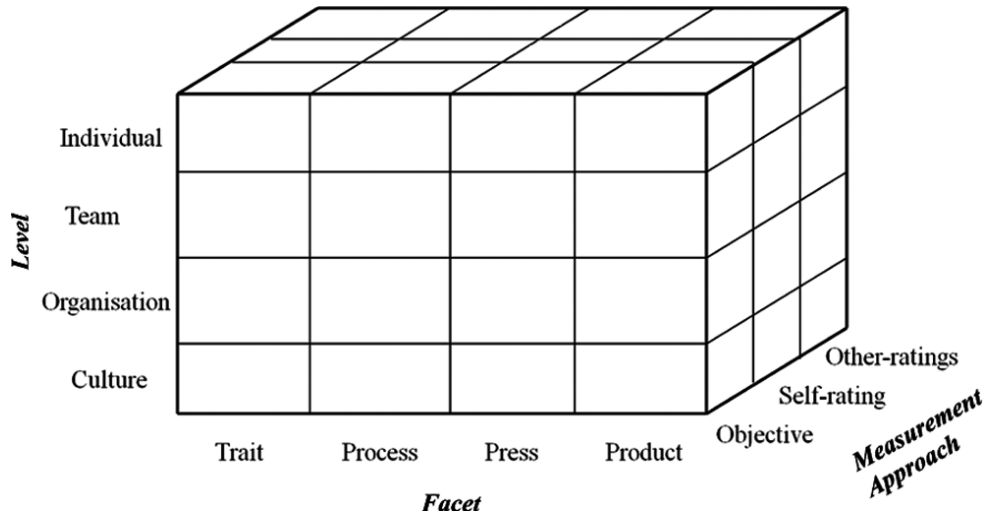


Figure 2 Measures of Creativity (Batey, 2012)

From a practice perspective, the assessment of creative ideas and output is especially important for organizations during the idea evaluation or "validation stage" (Montag et al., 2012; Csikszentmihalyi, 1996). However, organizational creativity research has shown managers do not differentiate between creativity, innovation, or problem solving, incorrectly viewing the three distinctly separate constructs as inextricably connected (Banks, Calvey, Owen, & Russell, 2002). Evaluating the most promising ideas and creative output during the iterative stages of the creative process requires accurate forecasting of which ideas and products are most creative, as well as which have the best chance for ultimate success (Daily & Mumford, 2006; Lonergan, Scott, and Mumford, 2004). Thus, evaluation of creative ideas and products is one of the most critical steps in the creative process, yet organizations struggle to make appropriate decisions about the creative ideas and products they develop (Harvey & Kou, 2013). For example, groups have been shown to prefer relatively average ideas produced by their members over more novel ideas created by other individuals (Rietzschel, Nijstad, & Stroebe, 2006). Unlike idea generation, which is considered a *divergent* creative thinking behavior, idea evaluation is a

convergent decision-making activity that occurs in later stages of the creative process and seeks to filter out poor ideas (Paletz & Schunn, 2010; Singh & Fleming, 2010). This convergent process of product or idea evaluation in group settings involves social, political, and critical assessment functions, as members "choose consciously or subconsciously to ignore ideas, advocate for their own ideas, show enthusiasm for others' ideas, and provide interpersonal rewards for good ideas" and "reflects the iterative and integrated nature of idea evaluation at the individual and organizational levels" (Rietzschel et al., 2006: 347). Organizations and individuals that are best able to discern which ideas and products are highly creative will not only produce the candidates that are most likely to be successful, but will also reduce the overall cost and timeframe of development.

From a research perspective, measurement of creative output is critical to the validity of experimental results. As noted above, various concerns have been raised about the evaluation of creative output in research; in particular the possibility of measurement misspecification as a result of differences between construct definitions and empirical measurement approaches (Sullivan & Ford, 2010). After a review of published creativity research in top journals over a 10-year period, Sullivan & Ford (2010) noted that, despite the growing use of definitions including both "novelty" *and* "usefulness" components, most creativity research studies continue to measure creativity as a unitary construct that relies on a single dimension. Other authors have suggested novelty and usefulness are associated with different organizational processes (Ford & Gioia, 2000), may have different causes and consequences (Reiter-Palmon, Illies, Cross, Buboltz, & Nimps, 2009), and may have distinct goals that can be represented as orthogonal constructs (Litchfield, 2008). As a result, some researchers hypothesize the differences between construct definitions and operational definitions in use may potentially create a misalignment

that threatens statistical conclusion validity and could result in errors of inference (Sullivan & Ford, 2010; Montag et al., 2012).

To test their hypotheses about possible measurement misspecification of prior creativity research, Sullivan & Ford (2010) conducted two experiments to examine alternative measurement models of creativity, comparing a one-factor measurement model with reflective measures, against two-factor and three-factor composite latent models using novelty and usefulness and novelty, usefulness, and 'stylistic appeal', respectively, as formative indicators. The results from both experiments suggest the two-factor composite latent construct model with distinct formative indicators for novelty and usefulness provided the best fit of the participants' assessments (Sullivan & Ford, 2010). However, the authors point out the measurement model required by a specific study may depend on the nature of research and its theoretical approach, and that in some circumstances a unidimensional or multiple-facet reflective indicator approach may be more appropriate. The authors conclude by noting a significant limitation of their research involved the use of students and professors to create and judge creative output, suggesting future research should investigate "the relative weight of novelty and usefulness assessments," as well as "how other appropriate judges evaluate the creativity of stimulus items in their domain" and "whether the depth or breadth of the judges' domain expertise affects creativity assessments" (Sullivan & Ford, 2010: 518). This research attempts to respond to the authors' three suggestions.

Other attempts to address the perceived insufficiency of the "novel and useful" construct and operational definitions have generally involved the addition of a new facet or criterion. As noted above, Sullivan & Ford (2010) included the criterion of "stylistic appeal" in their analysis in light of research suggesting "style" as another aspect of creativity, but did not find sufficient

support to unequivocally recommend adding that factor. Boden (2004) advocated three facets, requiring a creative idea to be novel, valuable, and surprising, following Bruner's (1962) definition of creativity as involving an element of "effective surprise." Simonton (2012) and others have advocated that researchers adopt a definition similar to the three-part test utilized by the United States Patent and Trademark Office (USPTO) for patent applications, under which a product or idea will only be awarded a patent if it is shown to be: 1) novel; 2) "non-obvious;" and, 3) useful. Section 101 of Title 35 of the United States Code governs the Patentability of a product or idea, which asserts:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Section 102 of Title 35 U.S.C. expands on the newness requirement of Section 101 by stating the invention must exhibit "novelty," which is defined as not having already been patented, described in a publication, or otherwise in public use, for sale or available to the public. This requirement is also described as the "prior art" exclusion and is designed to preclude obtaining a patent by appropriating someone else's previously published idea or publicly available product. Section 103 adds the second criterion of "non-obvious subject matter":

> A patent for a claimed invention may not be obtained…if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious … to a person having ordinary skill in the art to which the claimed invention pertains.

The "non-obviousness" criterion, or "surprise" in Simonton's (2012) formulation, similarly requires that a product or idea be sufficiently "above or beyond the current state of the

art" such that is not simply an obvious extension of an existing idea or concept (Barton, 2003: 476). European patent regulations contain very similar requirements of novelty and usefulness, but term the non-obviousness requirement as "inventiveness," requiring that the product or idea represent an "inventive step" or significant iteration in development either through the combination of disparate ideas or departure from a normal development cycle (Barton, 2003). Although some researchers have used a three-part criterion in the past, Simonton points out that "few have followed their example" (Simonton, 2012: 98, fn, 1).

II.3.1 *Consensual Assessment Technique*

Most studies of creativity that involve the assessment of creative products or output rely on external appraisals (Batey, 2012). The assessment of creative products by others initially began as "aesthetic judgment" of the arts more than 100 years ago (Kaufman, Baer, Cropley, Reiter-Palmon, & Sinnett, 2013). Amabile (1982) built on that historical approach in developing the Consensual Assessment Technique (CAT), which provides specific guidelines for measuring the results and output of creative effort. Amabile's consensual assessment technique is expressly premised on a "product-based" *operational* definition of creativity:

> A product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated. (Amabile, 1982: 1001.)

Amabile's CAT is unabashed in its reliance on knowledgeable individuals' subjective and intersubjective judgment of creative products, acknowledging the difficulty, if not near impossibility (and perhaps the irrelevance), of attempting to assess creativity objectively (Amabile, 1982, 1996). Moreover, CAT rests on two important assumptions: first, that given an appropriate group of judges, it is possible to obtain reliable judgments of creativity; and second,

that creativity exists on a continuum, i.e., some products and outcomes are more creative than others are. This first assumption takes into account the fact that, while it is difficult to identify specific features of a product that make it creative, people can recognize creativity when they see it. Second, as a *consensual* technique, CAT assumes there are degrees of creativity and that experts and individuals familiar with the domain can generally agree on the level of creativity of a particular product (Amabile, 1982, 1996). The consensual assessment technique follows the manner in which creativity is judged in many real-world situations in that it requires a group of judges, usually domain experts, working independently and without specific instructions, to make intersubjective assessments of creativity (Kaufman & Baer, 2012). As a result, CAT has been called the "gold standard" of creative output measurement for research (Kaufman & Baer, 2012).

Recent research highlights the importance of expertise in judging creative output using consensual assessment techniques and the stark differences between expert and novice ratings of creativity. For example, Kaufman, Baer, Cole, and Sexton (2008) assessed the level of interrater agreement among domain experts and college students who were asked to judge the creativity of poems and short stories. Novices were found incomparable with experts in both cases. For poetry, the correlation between experts and novices was only $r = .22$, and for short stories the correlation was $r = .71$ (Kaufman et al., 2008; Kaufman, Baer & Cole, 2009). The expert judges had very high levels of interrater reliability for both the poetry and short stories, with a coefficient α of .83 and .92, respectively. For novices, interrater reliability was surprisingly also very high, with a coefficient α of .94 for the poems and .93 for short stories. However, the very high interrater reliabilities of the novices required using scores from 106 novice judges. When any randomly selected group of 10 novice judge scores were analyzed, interrater reliability fell

dramatically to just .58 for poetry and .53 for short stories. Even the moderate level of agreement of r = .71 between novices and experts for short stories necessitated reliance on more than 100 novices. As a result, the authors concluded novices could not be relied upon to produce valid creativity judgments, and that the validity of CAT as a measurement tool in research requires the use of domain experts as judges (Kaufman et al., 2013).

Contrary to the evidence showing that domain expert judges are necessary for the consensual assessment technique to be considered reliable and valid, and that experts perform at a much higher level as judges than novices, a review of published research suggests the vast majority of creativity studies over the past 30 years utilizing CAT or similar consensual techniques to assess creative outcomes have not used recognized domain experts as judges (see Appendix 7.2, Summary of Published Creativity Research). The review reveals that a high preponderance of studies (37 of 41) used college students, research assistants, university professors, or study researchers to judge the creative outcomes produced by study participants. The use of non-expert judges to assess creative outcomes—where CAT necessitates expert judges for reliability and validity—raises questions about the soundness of those research results, as well as the appropriateness of the criteria and processes used to evaluate creativity in those studies (Sullivan & Ford, 2010). However, how experts actually judge creativity in practice is not well understood.

II.4 Experts and Expertise-Based Intuition

Psychologists have studied expertise for over a century in an attempt to understand how experts acquire and use knowledge, particularly studying the differences in mental functioning between experts and novices (Andre & Gobet, 2008). Current research on expertise suggests that two major phenomena constitute the foundation of cognitive expertise—visual perception and

knowledge organization. Visual perception plays a major role in evoking knowledge (Henderson & Hollingworth, 1999) and, because expert knowledge is fundamentally different from that of novices, the study of cognitive expertise cannot be separated from the study of perception—experts literally "see" their domain differently than novices (Reingold, Charness, Pomplun, & Stampe, 2001). Experts are also fundamentally different from novices in the way they acquire, manage, and organize domain knowledge (Andre & Gobet, 2008). While experts have not been shown to possess greater memory or computational capacity than others with similar intelligence (Ericsson & Ward, 2007; Ericsson, 2014; but see Grabner, 2014), the manner in which they develop and reorganize information sets them apart from novices of similar general abilities (Chase & Simon, 1973).

Expert knowledge is thought to be organized in two ways, as "chunks" or "scripts" at a low level of abstraction, and in more schematic and abstract elements or "templates" which can be adapted readily to a larger range of situations in higher levels (Andre & Gobet, 2008). Experts use these chunks and templates to situate problems or challenges within groups of past experiences from long-term memory, allowing the expert to extrapolate or anticipate next steps, the way an expert chess player can both "see" the entire chess board and simultaneously imagine most if not all of the next possible moves available (Gobet & Simon, 1996). Obtaining the requisite knowledge and schemas to become an expert requires a considerable number of years of direct experience and "deliberate practice" (Ericsson, Krampe, & Tesch-Romer, 1993).

Interest in the effectiveness of intuition in judgment and decision-making has also grown in recent years as scholars have begun to accept the concept, particularly with regard to expert use of intuition. Several researchers have suggested that expert intuition might be an effective way to manage the difficult trade-off between decision accuracy and speed (Dane & Pratt, 2007).

In psychology, intuition is generally accepted to arise from one of two information-processing systems that operate within the brain (Epstein, 2003). Under this dual-process model, decision making is hypothesized as either the product of primal, automatic, associative and experiential "System 1" processing, or the result of rational, intentional, deliberate, and extensional "System 2" processing (Epstein, 2003; Kahneman, 2003, 2011). Intuition is thought to arise from the faster, associative, and "nonconscious" information processing of System 1 (Epstein 2003; Kahneman 2003, 2011).

In management literature drawing upon dual-process models, the construct of intuition consists of four characteristics: "Intuition is a (1) nonconscious process (2) involving holistic associations (3) that are produced rapidly, which (4) result in affectively charged judgments" (Dane & Pratt 2007: 36). The process of "intuiting" is rapid (often instantaneous), spontaneous (without effort and unable to be controlled) and alogical (not necessarily contradicting the rules of logic but may not follow them either) (Dorfler & Ackermann, 2012).

Intuiting has been theorized to involve matching environmental stimuli or a non-algorithmic task with nonconscious patterns, schemas, or cues (Dane & Pratt, 2007; Klein, 1998) or linking or associating disparate elements of information (Epstein, Pacini, Denes-Raj, & Heier, 1996). However, a significant body of research shows intuition is ineffective in decision-making involving algorithmic tasks that can be decomposed and solved logically, sequentially, or mathematically (Dane, Rockmann & Pratt, 2012). In fact, research has repeatedly shown that reliance on intuition in decision-making that involves decomposable tasks, mathematical calculations, or following a sequence of rules is likely to produce gross misjudgments (Tversky & Kahneman, 1974; Kahneman, 2003). In their research on "heuristics and biases," Kahneman and Tversky define intuition as "thoughts and preferences that come to mind quickly and without

much reflection" (Kahneman 2003: 449). The results of heuristics and biases research highlighted the insufficiency and errors that arise from the use of any of three "intuitive" heuristics in decisions made in algorithmic tasks such as estimating the population of a city or choosing between limited alternatives presented: (1) the representativeness heuristic (i.e. "what is typical"); (2) the availability heuristic (i.e. "what comes easily to mind"); (3) adjustment and anchoring (i.e. "what happens to come first") (Kahneman 2003, 2011; Tversky & Kahneman, 1974).

However, other research suggests intuition may be much more effective and significantly less error-prone than analysis in cases of non-algorithmic and non-decomposable tasks (Hammond et al., 1987). Examples of non-decomposable tasks theorized to be more amenable to intuitive decision-making include judgments about artwork, the taste of food, and the morality of behavior (Dane et al., 2012; Haidt, 2001). Research into expertise-based intuition has focused on how experts recognize and retrieve large "chunks" of information, and use patterns or schemas stored in long-term memory without conscious effort when dealing with non-algorithmic situations (Dane & Pratt, 2007). Studies of the effective use of expert intuition indicate greater time pressure will cause an individual to rely more on intuition than analysis, and that experts possess highly sophisticated, nonconscious cognitive structures that afford rapid and accurate retrieval of information and creation of appropriate responses in time-critical situations (Klein, 1998; Simon & Chase, 1973; Dane & Pratt, 2007). However, research has also shown that not all expertise is the same, and expertise in one domain does not easily transfer to another (Kahneman, 2011).

As a result of these studies, intuitive judgments and decisions by domain experts at a very high level of expertise, variously called "intuition-as-expertise" (Sadler-Smith & Shefy, 2004),

"intuitive expertise" (Kahneman & Klein, 2009), and "expertise-based intuition" (Salas, Rosen, and Diaz-Granados, 2010), are theorized to be highly trustworthy, fast, and effective (Dane & Pratt, 2007).

> The distinctive earmarks of intuition are rapid response (a matter of seconds) and inability of the respondent to report a sequence of steps leading to the result—even denial of awareness of such steps . . . what impresses observers about intuition is that responses, especially those of experts, are frequently correct even though they seem to have required almost no processing time or effort" (March & Simon, 1993: 11).

Researchers have also concluded the most effective intuitive judgments are those of domain experts who have had the opportunity to acquire domain specific information and who have received critical feedback and cues from a high validity environment (Kahneman & Klein, 2009), and where the effectiveness of intuition relative to analysis is amplified by higher levels of domain expertise (Dane et al., 2012; Kahneman & Klein, 2009; Salas, et al., 2010).

II.5 Judgment and Decision Making (JDM)

For this research, understanding how experts use intuitive expertise to judge creative outcomes suggested application of two competing theories of judgment and decision-making. The two theories, Cognitive-Experiential Self-Theory (CEST) and Cognitive Continuum Theory (CCT), were applied in this study using Naturalistic Decision Making (NDM) models and techniques. An in-depth review of the literature on CEST and CCT helped shape the research conducted using NDM, and a summary of the literature is provided below.

II.5.1 *Cognitive-Experiential Self-Theory*

Cognitive-Experiential Self-Theory (CEST) is a dual-process theory of human information processing integrating theories of learning, cognition, personality, and the self

(Epstein, 2003; Norris & Epstein, 2011). CEST assumes the human brain consists of two complimentary but independent systems of information processing: an "experiential (i.e., intuitive) system" and a "rational (i.e., analytical) system" (Epstein et al., 1996). The experiential or intuitive system, sometimes referred to as System 1, is preconscious, fast, nonverbal and holistic, and is considered a product of human evolution operating automatically under principles of associative learning that solves problems through adaptation by reacting according to its experience and reinforcement history (Norris & Epstein, 2011). In contrast, the rational or analytical system, System 2, is a conscious, slow, affect-free, verbal reasoning system that is engaged deliberately to process information logically and solve problems through evaluation of evidence and alternatives (Norris & Epstein, 2011).

CEST posits the two processing systems often work in parallel and interact as necessary to contribute to a person's behavior and conscious thought (Akinci & Sadler-Smith, 2012). Because the experiential system is fast, requires little in cognitive resources, and is guided by prior experiences, it is particularly adept at handling the vast majority of information processing that occurs outside of conscious awareness on a daily basis (Epstein, 2003). This default to System 1 processing is thought to preserve cognitive resources and conscious deliberation for more abstract, logical, and challenging cognitive processing. The two systems are also presumed to respond differently to stimuli and use different cognitive resources. For example, when faced with an emotionally significant event, CEST predicts the experiential system will automatically search long-term memory stores for related "experiences" and emotional accompaniments that create an intuitive response (Akinci & Sadler-Smith, 2013). The rational system, on the other hand, is inferential, and operates through reason, abstract thought, and the use of language. Because of these criteria, the rational system requires more cognitive resources and has limited

processing capacity. Unlike the experiential system, the rational system is self-aware and has capacity to understand and decide whether to accept or reject influence from the experiential system. The rational system theoretically operates without emotion and can be changed through appeals to logic and reason (Epstein, 2003). As a result, even when operating under guidance from the experiential system, individuals are capable of discounting System 1 influence by consciously deciding to do so (Epstein et al., 1996). The theorized characteristics of System 1 and System 2 are shown in Table 1.

Table 1 Characteristics of Experiential-System 1 and Rational-System 2 (from Epstein, et al. 1996)

System 1 (experiential / intuitive)	***System 2 (analytical / reflective)***
Holistic	Analytic
Automatic	Intentional
Emotional	Logical
Mediates behavior by "feel"	Mediates behavior by conscious appraisal
Fast for immediate action	Slow for delayed action
Resistant to change	Easily changed through reason
Preconscious	Conscious

According to CEST, the experiential and rational systems tend to work independently but in tandem, "toggling" sequentially as needed, and affect one another depending on the nature of the stimulus (Epstein, 2003). Each system has its own unique adaptations, advantages, and disadvantages that tend to direct the order of operation of the two systems and the ultimate behavioral response. The experiential system adapts through implicit learning by experience, whereas the rational system adjusts through logical inference and explicit learning (Epstein 2003).

In operation, CEST predicts that when confronted with an event, the experiential system responds quickly and non-consciously by making automatic associations to similar past events or

experiences, eliciting a holistic emotional response or "vibe," which then directs further processing (Epstein, 2003). The slower, more analytic rational system subsequently attempts to understand or "rationalize" the initial emotional response and logically analyzes behavior before selecting the most emotionally satisfying and intellectually appropriate explanation for that behavior. "Such rationalization is a routine process that occurs far more often than is generally recognized," and the influences of the experiential system on the rational system are attributed by CEST "as major sources of human irrationality" (Epstein, 2003: 162). This may be because the experiential system is fast and its automatic emotional response is able to bias subsequent processing in the rational system outside of an individual's conscious awareness (Norris & Epstein, 2011). Likewise, the rational system has the ability, through conscious thought and deliberation, to correct the experiential system and teach it through repetition of thoughts and behaviors to adapt to new situations (Epstein 2003). However, the experiential system is thought to be much more efficient, able to manage greater amounts of information, and for that reason, tends to be the default processing system in daily life. "The rational system is capable of high levels of abstract reasoning and is therefore the source of humankind's unique accomplishments, but it is too effortful to efficiently direct most behavior in everyday life" (Norris & Epstein, 2011: 1044).

While dual-process theories, such as CEST, posit \ the two modes of cognition are often in direct conflict, each operating independently at different times, dual-process theories provide few, if any, specifics about how the two systems or modes interact or the nature of their relationship (Dhami & Thompson, 2012). This "either-or" approach to cognition as being either purely analytic or purely intuitive potentially creates a false dichotomy that has been challenged by competing views seeking to integrate analysis and intuition into a single theoretical

framework (Dhami & Thompson, 2012). However, recent studies of human brain functioning and mental processing using fMRI suggest cognition may actually operate in the manner predicted by dual-process theories. Researchers conducted fMRI studies of individuals during tasks designed to engage the two different brain networks: the task-negative or default mode network that operates outside of conscious attention (System 1), and the task-positive network that manages attention-demanding cognitive tasks (System 2) (Jack, et al, 2014). Results of analyzing brain scans of the subjects while engaged in the two types of tasks revealed that social tasks deactivated regions associated with mechanical reasoning, while mechanical tasks deactivated regions of the brain associated with social reasoning. The researchers concluded a physiological constraint on human processing might prevent both networks from acting simultaneously (Jack, et al, 2014).

II.5.2 *Cognitive Continuum Theory*

In contrast to CEST and other dual process models, Cognitive Continuum Theory (CCT) views different modes of cognition in decision making along a continuum, with intuition at one end and analysis at the other, and explains the interaction between different cognitive modes during a decision-making task (Dhami & Thompson, 2012). CCT, founded on social judgment theory and Brunswickian principles of functionalism, argues that cognitive processes and performance should be described and measured relative to the environment in which they function (Dhami, Hertwig and Hoffrage, 2004). Recognizing the benefits of both analysis and intuition, as well as the limitations of each to high-fidelity decision making, Brunswick introduced the concept of "quasirationality" as cognition that occurs when intuition and analysis are simulataneously present to some degree (Dhami & Thompson, 2012).

According to CCT, most decision making involves a combination of intuition and analysis, with cognitive tasks arranged along a continuum in terms of the task's tendency to induce intuition, quasirationality, or analysis. When performing a task, individuals move along the continuum in response to the properties of the task (Dhami & Thompson, 2012). CCT predicts initial task success will inhibit movement along the continuum, or preclude a change in cognitive mode, while initial task failure will stimulate movement and search along the continuum, potentially causing oscillation between the two modes until a decision is reached (Dhami & Thompson, 2012). CCT also predicts decision performance is contingent on correspondence between task properties and individual cognitive mode or abilities (Dhami & Thompson, 2012; Hammond, Hamm, Grassia & Pearson, 1987).

The different modes of cognition along the continuum are identified by unique sets of properties defining either intuition or analysis (Doherty & Kurz, 1996). Some of these defining properties and resulting impacts on processing are outlined in Table 2 below.

Table 2 Defining properties of intuition and analysis (from Dhami & Thompson, 2012)

Property	***Intuition***	***Analysis***
Consistency / reliability of judgments or cognitive control	Low	High
Awareness of cognitive activity	Low	High
Speed of cognitive activity	High	Low
Memory	Little encoding	Complex encoding
Metaphors used	Pictorial, qualitative	Verbal, quantitative
Information use	Flexible	Consistent
Confidence in judgments	Low	High
Errors in judgment	Many, but small and normally distributed	Few, but large and non-normally distributed

Quasirationality is defined as when any combination of these properties of intuition and analysis are used in judgment and exists along the continuum between the two polar extremes (Dhami & Thompson, 2012). In practice, quasirationality is thought to be the most prevalent mode of

cognition, such that “it is rare for any task to involve pure intuition or pure analysis” (Dhami & Thompson, 2012: 320).

Cognitive tasks along the continuum that are expected to induce either intuition or analysis can also be differentiated by their respective task properties, as shown in Table 3 below. The number, nature, and degree of task properties present dictate the cognitive mode that will be induced. Depending on the nature of the task, intermediate levels or a combination of properties that separately induce intuition and analysis will likely result in quasirationality (Dhami & Thompson, 2012).

Table 3 Properties of tasks that induce intuition and analysis (from Dhami & Thompson, 2012)

Task Properties	***Intuition***	***Analysis***
Familiarity with task	Familiar	Unfamiliar
Prior training / knowledge of task	None	Some
Amount of information	> 5 pieces of information	< 5 pieces of information
Information presentation order	Simultaneous	Sequential
Information presentation format	Pictorial	Quantitative
Inter-relation of information	Redundancy	Independent
Interpretation of information	Subjectively	Objectively
Number of response options	Many	Few
Time pressure	High	Low
Feedback available	Little / none	Cognitive feedback
Outcome knowledge	Available	Unavailable

Many of the defining properties of intuition and analysis, and the properties of the tasks likely to induce intuition, analysis, or quasirationality, have been supported by studies testing cognitive control (Dunwoody, Haarbauer, Mahan, Marino, & Tang, 2000), linearity of organizing principles (Hammond et al., 1987), and distribution of errors (Dunwoody et al., 2000; Hammond, et al., 1987). Confidence in the decision-making process has been shown to be higher than confidence in the actual judgment with analytic cognition, whereas confidence in method is lower than in the ultimate outcome with intuition (Dunwoody et al., 2000; Hammond et al., 1987).

There is also some evidence supporting CCT's prediction that different task properties induce different modes of cognition, particularly with regard to the number of cues, redundancy among cues, cue weights, availability of organizing principles and degree of non-linearity of

organization (Dunwoody et al., 2000; Hamm, 1988; Hammond et al., 1987). Cognitive mode has been shown to shift during a task depending on the perceived difficulty of the task and is dependent upon the tendency of the task to induce intuition (Hamm, 1988). Lastly, Hammond, et al. (1987) found empirical evidence to support CCT's prediction that achievement is dependent upon correspondence between task properties and the cognitive mode employed (Hammond et al., 1987; see also Dunwoody et al., 2000). Research also lends indirect support to many other assumptions of CCT about how task properties alter the mode of cognition listed in Table 2, including the ways experts use intuitive processing, the amount of information these individuals are able to cognitively process, and how high time pressure leads to use of more intuitive strategies (Dhami & Thompson, 2012).

Cognitive Continuum Theory has been used in several different research contexts to examine expert judgment, including management (Mahan, 1994), nursing (Cader, Campbell & Watson, 2005; Standing, 2008), engineering (Hammond et al., 1987), clinical decision making (Hamm, 1988), and retail products (Mathwicka, Malthotrab, & Rigdon, 2002). For example, Mahan (1994) applied CCT in a management context to examine the effects of task duration and task uncertainty on decision performance. Mahan found that stressors, such as short task duration and task uncertainty, induced a shift toward intuition but also resulted in a decrease in decision performance (Mahan, 1994).

II.5.3 *Application of CEST and CCT using Naturalistic Decision Making*

Most of the research applying CEST and CCT has involved experimental studies in laboratory conditions. The Naturalistic Decision Making (NDM) framework provides a method for applying CEST and CCT to understand how experts make decisions in cognitively challenging field settings. NDM arose from a need to take decision making out of the laboratory

and understand how decisions are made in the real world (Klein, 1997). The NDM framework is a descriptive model that focuses on cognitive functions of decision making in the field, particularly emphasizing situational awareness, recognition primed decisions, and sensemaking (Lipshitz, Klein, Orasanu & Salas, 2001).

One of the hallmarks of NDM is the Recognition-Primed Decision (RPD) model. RPD describes how people use past experiences and knowledge to develop and recognize cues, patterns, and relationships to make decisions quickly and effectively. The RPD model was developed using cognitive task analysis to understand how experienced fireground commanders make appropriate and effective decisions under conditions of extreme time pressure, high stakes, and uncertainty (Klein, Calderwood, & Macgregor, 1989; Lipshitz et al., 2001). Klein and colleagues hypothesized that fireground commanders, faced with extreme time pressure, would not be able to generate and select from a large number of decision options as predicted by models of rational decision making. Instead, it was anticipated that fireground commanders would be restricted to choosing between just two options, a favored decision and a comparison (Klein et al., 1989). However, after conducting in-depth interviews of experienced firefighters about their experiences with 156 highly challenging fireground incidents, researchers discovered the firefighters were not comparing any options; instead, they were generally following the very first course of action identified. This finding raised two questions: (1) how could the fire commanders effectively rely on the first option identified, and (2) how could they evaluate a single option without comparing it to any others? (Lipshitz et al. 2001).

In seeking answers to these questions, the fireground commander interviews lead researchers to identify three variations of the Recognition-Primed Decision model. In the first variation, a skilled decision maker sizes up the situation and develops a feasible course of action

using prototypes based on experience (Lipshitz et al., 2001). This variant works best in relatively straightforward situations under short time constraints. In the second variation, most often used when a situation is unclear or ambiguous, the skilled decision maker relies on a story-building strategy to mentally simulate the events leading up to the observed features of the situation, gauging the status of the situation to develop a feasible course of action based on the mental simulation (Pennington & Hastie, 1995). The third variation describes how fireground commanders are able to evaluate a course of action without comparing it to other options. Here, commanders develop and evaluate the first feasible course of action through mental simulations to determine whether the strategy might work and look for potential unintended consequences. The course of action then can be modified and re-simulated if it lacks "progressive deepening" until an acceptable choice is identified (Lipshitz et al., 2001).

All three variations of RPD rely heavily on the domain expertise and experience of the decision maker, and domain expertise greatly enhances the speed and quality of judgments. In the first variant of RPD, expertise provides the prototypes the decision maker uses to quickly categorize the situation and recognize a possible course of action. In the second and third variants, expertise is used to develop the stories, mental models, and simulations required to imagine the events leading up to the situation, develop a feasible course of action, and simulate and test the possible outcomes if that course of action is taken (Lipshitz et al., 2001). Expertise in the RPD model also allows decision makers to respond quickly to changing conditions and ill-defined goals because RPD focuses on working forward from existing conditions: "Therefore, the RPD model is a blend of intuition and analysis. The pattern matching is the intuitive part, and the mental simulation is the conscious, deliberate, and analytical part." (Klein, 2008: 458).

Research has also shown that experts are more likely to use "forward-chain reasoning,"

whereas novices and non-experts tend to rely on "backward-chain reasoning" (Patel & Groen, 1986). Forward-chain reasoning involves taking the known or assumed aspects of an event or series of events and imagining next steps and ultimate conclusions. Forward chaining requires significant experience and understanding of the possible consequences of action or non-action and the ability to discover missing elements. Forward chaining involves "If-Then" reasoning where "If" is the known and "Then" relates to the inferred effects. Backward-chain reasoning, on the other hand, involves starting from a known result or consequence and developing hypotheses about causation and linkages to see if necessary conditions are met to achieve the desired outcome (Mess, 1995; Darden 2002).

The three variations of RPD show how expert decision makers can be effective in the face of uncertainty, time pressure, shifting conditions, and ill-defined goals without developing and comparing multiple choices or following a rational choice approach (Lipshitz et al., 2001). These findings have been replicated by research of experts in various military, design engineering, offshore oil industry, commercial aviation, and medical settings (see Klein, 1998 for a review). However, this additional research also suggests several boundary conditions: RPD tends to apply in situations where the decision maker has sufficient relevant domain experience and expertise, is under significant time pressure, and where ill-defined goals produce uncertainty. RPD is less effective in situations presenting highly combinatorial or algorithmic problems, where justifications are required, or in cases where differing views of multiple stakeholders must be considered (Klein, 1998).

As a descriptive model, NDM uses a variety of field observation techniques to understand real-world decision making within the context of the task environment to gain insight into sources of difficulty, error, non-optimal performance, and how larger systems support or inhibit

decision makers in practice (Lipshitz et al., 2001). NDM methods used for eliciting decision-making strategies and expert knowledge include: structured and unstructured interviews, critical incident analysis, domain and concept maps, think-aloud protocols, simulations, and real-time observation or review of recorded decision-making behaviors (Klein et al., 1989; Lipshitz et al., 2001). One of the key techniques of NDM is *cognitive task analysis* (CTA), a broad term that encompasses multiple methods for capturing experts' and practitioners' knowledge and processes in performing their jobs (Klein et al., 1989; Lipshitz et al., 2001). CTA and related techniques have been shown to provide reliable and valid results in a number of different studies (see Hoffman et al., 1998, for review).

For the purposes of this research, NDM was used as the framework for investigating how creativity domain experts make judgments of creative output "in the wild" (Lipshitz et al., 2001: 346). Expert judgment (assessment and evaluation) of creativity occurs in numerous real-world settings including education (e.g., art, design, creative writing), the workplace (e.g., graphic design, advertising, marketing, architecture, engineering), and society (e.g., museums, film, music, photography). CEST and CCT suggest experts may use many different approaches, techniques, and strategies in decision making, depending on the nature of the task and the surrounding environment. NDM lends a methodology to help uncover and understand the processes, cues, cognitive schemas and approaches that creative experts use when deciding if a product or idea is indeed creative.

Accordingly, this study utilized Applied Cognitive Task Analysis (ACTA), a streamlined version of cognitive task analysis developed for researchers untrained in cognitive psychology (Militello & Hutton, 1998; Crandell, Klein, & Hoffman, 2006). ACTA provides a refined set of cognitive task analysis techniques specifically designed to identify the key cognitive elements

required to perform mentally challenging or complex tasks (McAndrew & Gore, 2013). Because ACTA focuses on the key cognitive elements underlying difficult judgments and decisions, critical cues and patterns and problem-solving strategies, it is particularly well suited to studying expert decision making (McAndrew & Gore, 2013). A more complete description of how ACTA was used in this study is provided at Appendix 7.1

III METHODOLOGY

III.1 Introduction and Overview

The purpose of this multi-case field study was to investigate the processes and measures that expert judges use in assessing the output creativity of individuals who regularly generate products and ideas for industry. Understanding how individual domain experts make judgments about the creativity of products and ideas under extreme time pressure, uncertainty, and ambiguity is expected to provide useful insights into the criteria, weightings, key elements, and biases involved in judging creative output. These insights are also anticipated to advance understanding of the challenges creativity researchers face in assessing creative output, and to suggest methods and practices that will improve the measurement, reliability, and validity of this area of research.

To better understand this phenomenon, this study investigates how domain experts judge the creativity of entries in professional awards programs through the following specific research questions:

1. What criteria do expert judges of creativity awards contests report using to assess the creativity of products and ideas; in particular, do experts use the "novel and useful" definition of creativity that is employed in most creativity research or some other criteria?

2. What processes do experts use as they judge creativity in contest settings; are the processes involved in creativity contests similar to or different from those used in the experts' professional workplace settings?

3. What types of cognitive processing are involved in expert judgment of creative products and ideas; do experts use intuition, rational analysis, a combination of both, or something else when judging creativity contests?

4. What do experts say are the differences between how experts and non-experts make judgments of creativity; what kinds of mistakes might a novice make in judging creativity?

5. What biases and heuristics do expert judges acknowledge encountering when judging creativity in real-world situations, and how do they attempt to deal with them?

This chapter describes the study's methodological approach and explains: (1) the rationale for the research approach; (2) a description of the research participants; (3) a summary of the information required for the research; (4) an overview of the research design; (5) the methods used for data collection; (6) the methods used to analyze and synthesize the data; (7) ethical concerns; (8) issues of trustworthiness; and (9) limitations and delimitations of the study.

III.2 Rationale for the Research Approach

III.2.1 *Rationale for a qualitative field study approach*

This research investigated how domain experts make real-world judgments about the creativity of products and ideas. As such, it involved a process study of the steps, procedures, and actions judges take in evaluating creative output, and a cognitive behavioral study of the criteria, cues, and mental processing that expert judges use to form judgments. The focus of the research was on "how" particular individuals understood, experienced, and applied certain social constructs, and "what" information and approaches those individuals relied upon to make decisions in a social environment. Accordingly, a qualitative research approach was determined to be most appropriate for this field study.

Qualitative research is concerned with the social complexities of the real world in a particular context, over a specific period of time, and the ways in which participants make sense of the world around them (Merriam, 2009). As such, it derives from a constructivist epistemology that contends knowledge is a human and social construction based on

interpretations of available information (Lincoln & Guba, 2000). Qualitative research involves researcher participation to investigate social situations and interactions in an effort to achieve a more holistic understanding of the events, participants, and environments (Denzin & Lincoln, 2011). In contrast to quantitative analysis, which is more concerned with the relationships between variables, measurement of objectively verifiable data, and testing of hypotheses, qualitative research is a naturalist approach that seeks to extract and interpret the meaning of events as they are experienced and understood by individuals (Merriam, 2009). While quantitative research assumes a post-positivist view that an approximation of reality can be observed and described using evidence-based probabilities, qualitative research is based on an interpretivist perspective, wherein there are multiple versions of reality, requiring deeper inquiry of shared meaning from various sources (Denzin & Lincoln, 2011). In this respect, "model-dependent realism" argues there is no observer-independent concept of the world and reality:

> Model-dependent realism is based on the idea that our brains interpret the input from our sensory organs by making a model of the world. When such a model is successful at explaining events, we tend to attribute to it, and to the elements and concepts that constitute it, the quality of reality or absolute truth…According to model-dependent realism, it is pointless to ask whether a model is real, only whether it agrees with observation. If there are two models that both agree with observation…then one cannot say that one is more real than another. (Hawking & Mlodinow, 2010: 7 & 46).

III.2.2 *Rationale for a case study approach*

This field study was designed to understand "how" judgment and decision making occurs in a particular context, and as such, it was necessary to examine contemporary events that the researcher could not manipulate or modify by intervention. The information sought in this

research related to the actions, mental processes, choices and biases of individuals engaged in one particular complex activity. Thus, a case study approach was chosen as the most likely to yield the desired level of detail and understanding (Yin, 2009). Case studies maximize face validity by incorporating a contemporary problem that other researchers or organizations identify with, which allows the researcher to explore challenging questions in the context of complex and uncertain situations. However, case studies are also susceptible to potential disadvantages, including difficulty accessing participants and data, lack of control over the context, and significant time investments (Myers, 2009).

The study sought to investigate judgment and decision making at an individual unit of analysis by observing, interviewing, and learning from individual domain experts. As Myers (2009) explains, qualitative research aims to understand decisions and action in context by seeing and talking with people who are or who have engaged in the actions and events of interest. To increase the relevance of this study to practice, a pluralistic methodology of engaged scholarship was used (Van de Ven, 2007), which involved the participation of business stakeholders to uncover the complex, real-world problems they routinely face (Van de Ven, 2007). The study included perspectives and feedback from key stakeholders and participants in creative industries, such as domain experts, experienced judges, creativity awards program operators, managers of creative employees, and experienced creative employees throughout the research process.

Following an engaged scholarship approach, this research study adopted the seven principles recommended by Klein and Myers (1999) for conducting interpretive field studies: (1) the fundamental principle of the hermeneutic circle ("understanding is achieved by iterating between the interdependent meaning of parts and the whole they form"); (2) contextualization ("critical reflection of the social and historical background of the research setting"); (3) interaction

between researcher and subjects (critical reflection of how the data obtained were socially constructed); (4) abstraction and generalization (relating the data and observation to theory); (5) dialogical reasoning (awareness of possible contradictions between theoretical preconceptions and actual findings); (6) multiple interpretations (possibility of differences in participants' interpretations); and, (7) suspicion (awareness of participants' possible of biases and "distortion of events") (Klein & Myers, 1999). These principles are not all mandatory or required to be used in a particular order—researchers are expected to use their judgment and discretion in deciding how and when various principles should be applied while keeping in mind the principles are somewhat interdependent (Klein & Myers, 1999).

III.3 Research Setting and Context

This study is situated in a field setting to understand how individuals make judgments of creative products and ideas in the real world. In an effort to explain how judgments are made at a very high level, the inquiry focused on individuals with substantial domain experience in creative industries that were recognized as experts by their peers. In addition, it was important that study participants possess a significant level of past experience in making judgments about creative products and ideas, both in workplace settings and in situations involving time pressure, high stakes, and uncertainty.

Creativity award program judges are selected because their peers have recognized them as creative industry experts, and because they have received awards for their creative work in the past. These judges are also usually managers or directors in highly creative marketing, advertising, graphic design, photography, product design and packaging, architecture, and research and development industries. As such, they have experience evaluating creative products

and ideas at each step of the creative process in their capacities at work and in contests under stress, ambiguity, and time constraints.

After investigating of a number of creativity award programs, one program on the U.S. East Coast was selected as most appropriate for this study. All of the judges involved in the selected awards program were previous winners of multiple creativity awards, had more than 15 years professional experience in a creative industry, and were nominated as judges by other professionals. The awards program's primary purpose was recognizing and awarding "creativity" in a large number of professional fields, in contrast to many other programs where creativity comes second to other content areas, such as "advertising." The judges chosen for this program came from diverse backgrounds, industries, and major metropolitan cities in the United States, Mexico, Italy, Australia, France, the United Kingdom, Brazil, China, Singapore, and Spain. The awards program was also chosen because it is one of few that requires judges to be physically present to score items simultaneously in a large venue over several days, instead of viewing entries either via webpages or on a computer file. These conditions provided the judges the advantage of handling each item personally, interacting with other judges, and evaluating at their own pace, but also created a field condition of significant stress, compressed time constraints, uncertainty, and high stakes that are key elements for understanding how experts assess creativity. The setting also allowed the researcher to examine judges during actual contest conditions, and conduct simulations and think-aloud exercises with contest entries.

III.3.1 *Overview of information needed*

Obtaining information from diverse sources was considered important to investigate how expert judges assess creativity in field settings. To understand the tasks and cognitive challenges judges might face, it was critical to interview judges with a very high level of experience in a

creative domain, but also to observe them in actual field conditions. Furthermore, while semi-structured interviews provided valuable insight into participants' subconscious awareness of their actions, it was necessary to use a form of cognitive task analysis that involved simulation exercises and think-aloud protocols. Collateral written materials from the awards program, as well as interviews of the operators of the contest, were obtained to provide context about the judging event. Lastly, data from the judges' entry scores from this contest and from previous award contests over a five-year period were obtained in order to evaluate the inter-rater agreement among these expert judges.

III.4 Research Participants and Data Sources

III.4.1 *Pilot Study*

This study was inspired by challenges the researcher encountered in his professional experience as an executive at an architecture and design organization, as well feedback of other business executives about the difficulty of identifying highly creative work. These challenges also inspired the researcher and two colleagues to conduct an unpublished student research study examining the impact of rewards and task description on employee motivation and creative outcomes. As part of that initial study, the research team interviewed managers of several different organizations, each with more than 10 years of experience supervising creative employees and producing creative ideas and products themselves.

Using the information obtained through the interviews, a pilot study that featured a logo design contest for a charity organization was developed. Freelance graphic designers were asked to design a logo for a specified charity, and 36 different logo designs were submitted. Three independent graphics professionals scored the submissions separately based on the criteria of "novelty" and "usefulness," each on a scale of 1 to 5. During the scoring process, two of the

professionals residing in the U.S. scored one particular entry as highly novel and also very useful. However, the third professional, based in the U.K., scored the same entry as a 1 on both novelty and usefulness. When the third professional was asked why he scored the entry so low, he indicated the logo design appeared to be a very close replica of a finalist logo for the 2012 London Olympics. When the other two judges were asked about their scores on the same item, they revealed they had not seen the finalists for the London Olympics logo. To the U.S. judges, the entry appeared to be unique and to be a good fit with the charity's mission. This raised the issue of whether the criteria of "novelty" and "usefulness" are good measures of creativity where the full domain of existing art cannot be known by any one evaluator, and how (and in fact whether) these two criteria are assessed in real world settings.

The pilot study uncovered numerous other practical and theoretical challenges surrounding the discernment of creative outcomes in professional settings and academic research. Of particular note was the lack of understanding about how the creativity of professional work product is identified and measured, how those who assess creativity go about the critical task of selecting the most creative ideas and products, and the factors and criteria judges actually use to identify highly creative outcomes. These practical and theoretical questions contributed to development of the research questions for this study.

III.4.2 *Criteria for Research Participants*

To better understand the processes and cognition experts use when evaluating creativity, this study sought to identify research participants who were recognized by their peers as domain experts in various creative endeavors and industries. Of particular interest were individuals with more than a decade of substantial experience in creative industries who also had significant experience assessing creativity in professional settings and in judging creativity awards

programs. Numerous studies have shown that superior expertise is developed through repeated effort and diligent practice over a significant number of years in a specific domain environment that offers regular feedback and opportunity for learning (Ericsson, Krampe, & Tesch-Romer, 1993). In professional settings, experts are recognized by others in the industry through their years of experience, professional commitment, and dedicated practice, and as a result are often selected by their peers to acts as mentors, judges, managers, leaders, and educators.

Awards programs that recognize highly creative work in fields such as graphic design, advertising, and product development utilize such peer-recognized industry experts as contest judges to choose winners and increase the overall credibility of award selection. Judges for these types of awards programs are considered among the very best in their field and have often received numerous awards for their own work in previous contests. Thus, recognized industry experts who have judged award programs specifically focused on creativity and creative professions are anticipated to have established the level and type of experience this study seeks to understand.

III.4.3 *Participant Selection*

For this study, criterion-based and purposive selection methodologies were used to identify and select research participants with the requisite professional experience, skill, and expertise (Patton, 1990; Merriam, 2009). Purposive sampling allows for a selection of information-rich cases at the heart of the research investigation. Criterion-based methods require that all participants meet one or more criteria predetermined by the researcher, and that they have extensive experience in the identified phenomena to be investigated, in an effort to reduce variation (Patton, 1990). Internet research revealed more than a dozen different national and international award programs in various creative industries operating within the United States

between 2010 and 2015. This list was narrowed to four whose primary missions were to recognize and award highly creative work of professionals across a broad range of categories that also utilized recognized industry experts as contest judges. Specialty and niche award programs were excluded from further consideration if they focused on only one or a few categories within a defined creative industry. Also excluded from the list were those that did not identify creativity as a specific award criteria, those focused on technical innovations, and those that were aimed mainly at recognizing other aspects of professional work, such as overall financial performance or advertising effectiveness (e.g., the "Effie Awards", see www.effie.org). Award programs also were excluded if all of the judges or the jury for the program did not appear to be industry experts or otherwise have significant direct professional experience in the domain that was evaluated (e.g., general interest publication awards that utilized writers and editors from the publication as judges for technical product and other non-writing awards).

Program directors or lead judges of the four remaining creative industry award programs were contacted by email using information listed on the websites for each contest. All four identified program posted the identities of the judges from the most recent contest as well as the identity and contact information for the program director or judge in charge. An introductory email was sent to each of the four programs identifying the nature of the proposed research and inquiring about interviewing their program's judges. Follow-up emails were sent if a response had not been received within a week of initial contact. After two weeks, program directors from two of the four contests had responded to the inquiry, and both offered to discuss the research. After discussing the proposed research by phone, one of the two programs was selected as the most appropriate and willing to participate in the research effort. The selected program and the

program director have requested that their identity not be revealed in publication of the research study.

The awards program that served as the context for this study has held contests for more than 30 years, with entries received from around the globe, and more than 60 categories of creativity were assessed each year. In 2014, the awards program received more than 800 separate entries and gave awards in 64 distinct categories, including awards in print categories such as graphic design, product packaging, magazine and periodical design, newspaper and print advertising, annual reports, and book covers, as well as for digital and broadcast media design including website, web and smartphone applications, animation, film, training videos, and television and radio advertising.

III.4.4 *Backgrounds of the Participating Expert Judges*

All of the judges who agreed to participate in this research study had broad and diverse backgrounds in various creative industries, as well as having significant experience working in and managing creative organizations. All participants were working professionals with 15 or more years of direct experience in creative professions such as advertising, branding, product design, photography, copywriting, packaging design, digital media (web and application design), graphic design, and broadcast media. Several of the participants had experience in multiple fields, and some had a total of more than 30 years of experience in creative industry. Nine of the participants were male and three were female. Seven of the participants owned and operated their own design, branding, or advertising companies or agencies at the time of the study, and the remainder held positions of senior creative director, senior vice president, or managing director in their respective organizations.

Each of the study participants had been awarded numerous individual creativity awards for own their design work from national and international competitions including Best of Show, Best in Category, and Gold awards from various contests including the Cannes Lions International Festival of Creativity, the International Academy of Digital Arts and Sciences ("Webby's"), the American Advertising Federation ("ADDY's"), Communication Arts ("CA"), the American Graphic Design and Advertising Awards ("AGDA"), Creativity International Awards, and the Summit Creativity Awards. As past award winners, all of the participants had been invited and had acted as judges in multiple international, national, and regional awards programs, and many had served as a judge for more than 20 contests. Many are also recognized public speakers on creativity and design, and some had published best-selling critically acclaimed books on advertising, photography, and branding. The participant group for this study also included a past winner of the U.S. Presidential Design Award, a National Press Photographer of the Year, and a Pulitzer-prize winning photographer.

III.5 Data Collection

This research follows Yin's (2009) recommendations for data collection: (1) use of multiple sources of evidence; (2) development of a case study database; and (3) a documented chain of evidence (Yin, 2009: 114-124). The research included formal, semi-structured and probe question-based interviews using Applied Cognitive Task Analysis techniques to question domain expert judges. Appendix 7.1 provides a summary of the Applied Cognitive Task Analysis approach used in this study. Advice, feedback, and contextual information were also obtained from several creativity award programs through informal discussions with program directors, operators, and participants. Archival documents, including information from various creativity award program websites, presentations, contest rules and criteria, were also reviewed.

III.5.1 *Observation of Creativity Contest Judging Event, Simulations, Think-Aloud Exercises and In-Person Interviews of Experts and Contest Administrators*

In an effort to further understand the context, background, and activities of judging creativity in an awards show, the researcher obtained permission to attend and observe an awards contest-judging event in a major city in the eastern U.S. The judging event is held annually and takes place over a three-day period in a large hotel ballroom. Judges for each annual contest are selected by contest administrators based on recommendations from past judges, industry leaders, prior judging experience for this and other contests, industry experience and recognition as industry experts. An awards program director and administrative assistant administered the judging event.

Before the judging process began, the program director for the awards program introduced the researcher to the eight judges in a group setting, and allowed the researcher to give an overview of the research and offered judges the opportunity to participate in the study. The program director and administrative assistant also agreed to be interviewed for the study, and the researcher was allowed to observe the judging process over the full three-day event. During the course of the judging event, five judges agreed to be formally interviewed during breaks in the judging process. All eight judges and the awards program director and administrator also participated in general roundtable discussions about the awards program and overall judging process with the researcher after all the judging had concluded. In addition to the interviews and roundtable discussions, the five judges who consented to be interviewed also agreed to participate in think-aloud exercises as they scored a sample of actual entries or participated in simulated judging exercises using entries from the contest that had been previously scored.

The researcher observed all eight judges as they viewed, evaluated, and scored the contest entries over a three-day period. The formal scoring process began with a short introduction by

the contest administrator, an overview of the entry categories and explanation of how the entries were displayed in the ballroom, followed by instructions on how to use the electronic scoring devices and a review of the schedule of the three-day event. Each judge was assigned an electronic tablet with a program that allowed him or her to record a numerical score for each entry by category and to record nominations for the "Best in Show" award, which recognizes the top entries regardless of category.

Judges were told that many entries were entered in multiple categories; for example, a photograph in a magazine advertisement might be entered in both the advertising photography category and the magazine photography category. Each judge was also given a one-page explanation of the instructions on how to the use the tablet. Included on this sheet was a list of criteria for the judge to use in scoring entries and an explanation of how awards would be selected:

> **Judging Criteria**
>
> - 9-10 being excellent! Top of its category ("I wish I had done that!").
> - 7-8 being above average work for the category.
> - 5-6 being average work for the category.
> - 3-4 being below average work for the category.
> - 1-2 should not be considered for an award.
>
> Scores from all judges will be averaged to give the final score. Best in Show candidate selections will be pulled by staff at the end of the judging and all nominations will be discussed by the Jury on Saturday.

The judges received no other instructions or directions on how to judge the entries, how to go about reviewing the categories, or what was expected of them.

Judges were observed over the three days as they moved around the room from table to table where entries were displayed, and the researcher kept notes of the judges' actions and

behaviors. In some instances, the researcher surreptitiously timed (with permission) the participant judges as they reviewed individual categories and entries, in an attempt to capture how long participants spent forming assessments. The researcher also took notes as to whether and when participants worked alone or in concert with other judges, whether judges spoke to each other about individual entries, and what judges said aloud about the entries and process. During breaks in the scoring, judges and the contest administrators were observed interacting and, on several occasions, the researcher engaged in conversation with the group about the contest, the overall quality of the entries, the research study, and the scoring process. At different points during lunch and other long breaks, each of the participants was interviewed in a location away from the other judges. Think-aloud scoring was conducted during judging sessions, and simulation exercises were held during breaks in the scoring process.

Interviews with the participants during the judging event focused on how each individual approached the task of judging creative products and ideas, their judging experiences, the criteria used to assess contest entries, the challenges of judging creativity in a contest setting, the differences between assessing creativity for an award and in professional practice, the types of heuristics and techniques participants used to assess creativity, and the potential biases participants encountered and considered during the assessment of creativity. Participants explained that the awards program paid for the judges' accommodations and provided meals during the event; however, none of the judges were compensated for their participation and each had to pay for their own travel. During a round-table discussion after the event, many of the judges said they participated in judging events out of a sense of obligation to the industry, as a way of staying connected to the very best work in their profession, and to strengthen their professional credentials.

After participants had explained as much as they could recall about the assessment process, the interviewer followed the Applied Cognitive Task Analysis protocol described in Appendix 7.1 to delve into the cognitive aspects of the judging task. Participants were first told to "think about what you do when you score contest entries" and asked to break down the task of assessing creative products or ideas into three-to-six steps, and then identify which of the steps required the most cognitive effort. The interviewer then used the elicited steps to create a "task diagram" that provided a broad overview of the judging task and was reviewed with the participants for accuracy and completeness. Generic and specific probe questions were then asked as part of the "knowledge audit" step to elicit greater detail about each of the identified subtasks and the cognitive effort necessary to complete them effectively, including critical cues and strategies for judging creativity broadly and scoring each entry specifically. Concrete examples and specific information about past experiences and comparisons between how experts and novices might perform the task were captured during this stage of the interview to create an inventory of task-specific expertise. Participants identified specific cues and strategies that were explored further to identify how they were used to make decisions, and to help explain why the task poses challenges to inexperienced or novice judges. During the simulation interview, five participants were provided with actual contest entries from a selected category and asked to score each entry while describing aloud the processes and cognitive steps they undertook. In three cases, the participant had not judged the entries beforehand, such that their descriptions and scoring were observed in real-time. In the other two cases, participants had scored the entries previously, and they repeated their assessments during the interview, describing their thinking and process aloud as each entry was viewed again.

The simulation interview was used to build on the information obtained in the first two steps to contextualize the task, allowing the interviewer to better understand expert judges' cognitive processes. After the interviews were complete, a "cognitive demands table" of the information elicited in all the interviews was created to merge and synthesize the data. The cognitive demands table provided a format for the researcher to identify those areas that required complex cognitive skills and pertinent problem-solving and decision-making activities involved in the task of assessing creativity.

III.5.2 *Interviews of Judges from Prior Year Contests*

After the awards contest judging event, additional participants for this study were solicited by email from a list of past judges provided by the awards program director. The list contained names and contact information for 18 judges who had participated in at least one of three creativity awards contests preceding the 2014 program. The solicitation email described the research in general terms and requested a response if individuals were interested in participating in the research. Three past judges responded and agreed to participate within the first week of the initial email. A second email was sent to the remaining judges on the list the following week, and two more subsequently agreed to be interviewed. A third and final email was sent two weeks later, at which point, two more judges agreed to participate. Two judges declined due to busy schedules. The remaining nine judges on the award program's list did not respond to any of the three emails sent.

Interviews of the seven judges who agreed to participate were conducted by telephone following a semi-structured approach. Each interview lasted between 60 and 90 minutes. Interview questions were drawn from those posed to judges during in-person interviews at the contest event. The questioning focused on how the participants generally approached the task of

judging creative products and ideas, their judging experiences, the criteria used to assess contest entries, the challenges of judging creativity in a contest setting, the differences between assessing creativity for an awards program and in professional practice, the types of heuristics participants used to assess creativity quickly, and the potential biases participants were aware of during the assessment of creativity. Following the Applied Cognitive Task Analysis protocol, participants interviewed by phone were also asked to break down the judging task into steps to identify those that required the most cognitive effort, which helped develop a task diagram. Probe questions were then asked during the knowledge audit as to subtasks and cognitive effort, critical cues and strategies employed. Past experiences and comparisons between experts and novices were explored to create a task-specific inventory with examples. For the simulation portion of the telephone interviews, participants were presented with a challenging brief scenario drawn from the in-person judging think-aloud exercises that had been designed in advance specifically for this purpose. Participants were asked to imagine the scenario, visualizing each step involved, and describe their thinking and processes. The participants were then probed on issues relating to situation assessment, potential errors and biases, cues and patterns, and other challenges the situation might present. The information obtained from these simulation interviews was added to the cognitive demands table for further analysis and synthesis with data from the in-person interviews.

III.5.3 *Collection of Scoring Data from Observed Event and Prior Contests*

After the awards contest judging event concluded, all awards were announced, and all interviews had been completed, numerical scoring data from the observed contest and nine prior contests were obtained from the award contest program administrator. The dataset received assigned a unique anonymous number to each judge and a unique randomly assigned number to

each entry. Data for ten contests was provided, however, due to apparent errors in the random assignment of some item numbers, only scores from eight contests were suitable for analysis. Over the eight contests analyzed, 70 judges scored a total of 8,699 entries. According to the program director, not every item was scored by every judge in each contest for various reasons including lack of sufficient time, insufficient familiarity with a particular contest category, and judges' unwillingness to rate specific entries with which they were already familiar. Thus, the number of scores analyzed across the eight contests totaled 63,809.

III.6 Data Analysis

This study follows Miles and Huberman's (1994) three recommendations for qualitative case data analysis: (1) data reduction, (2) data display, and (3) conclusion drawing and verification (Miles & Huberman 1994: 10-12). The data analysis and collection process was interactive and cyclical, allowing the researcher to reflect, revise, and reconsider, gaining a deeper understanding of the information collected.

III.6.1 *Data Reduction*

Data reduction is a process that involves selecting, focusing, simplifying, abstracting, and transforming collected data into more usable forms (Miles and Huberman, 1994). Data reduction occurred continuously throughout this study from before the project began until the final report was completed. To improve validity and assist in analysis, this research included the use of various methods for summarizing (contact summary sheets, document summaries, case analysis, and interim case summaries), different approaches to coding (at both descriptive and inferential levels), methods to assist in thinking about data (annotations, journaling, and memos), and methods for producing extended reports throughout the study.

III.6.2 *Data Display*

Miles and Huberman (1994) also recommend the use of matrices, graphs, charts, and diagrams designed to assemble organized information into immediately accessible, compact forms. Creating a data display was an iterative process, occurring throughout and following data collection. Additionally, displays were created to compress and order data to allow the researcher to draw justifiable conclusions. Single-case displays were collected into a matrix, which were further condensed to permit side-by-side comparisons (Miles & Huberman, 1994: 176). Cross-case displays, composite models, and sequence analysis were particularly helpful in data analysis by allowing for ordering actions and behaviors and sorting by significant categories (Miles & Huberman, 1994: 172-187).

III.6.3 *Conclusion Drawing and Verification*

Data reduction and data display provided the base information for conclusion drawing and verification. Review of the data collected involved identifying key themes and drawing conclusions by identifying cues, patterns, processes, justifications, and explanations from the obtained and observed data. Initial conclusions were kept loose and tentative until further support was gathered and solidified as the process concluded. The conclusions were then verified after analysis to improve validity.

III.6.4 *Coding*

All interviews were recorded and transcribed by the researcher or a professional transcriptionist. Following the recommendations of Miles and Huberman (1994), data was coded using both descriptive and inferential codes to facilitate analysis and interpretation. An initial coding scheme for the interview transcripts of expert judges was developed based on a review of literature of various definitions and components of creativity, the measurement and assessment

of creative output, dual process models of judgment and decision making, the various elements and indicators of expertise and expert cognition, the factors and differentiators of Cognitive Experiential-Self Theory and Cognitive Continuum Theory, indicators of heuristics and potential biases, and the aspects of Naturalistic Decision Making, including three different forms of the Recognition Primed Decision model. From the literature review, 11 major codes and 32 minor or sub-codes were developed for use in the review and coding of interview transcripts. The review and coding process inductively generated 3 additional major codes and 14 additional sub-codes as several unique concepts, descriptors, and unanticipated aspects of the judging criteria and process arose from explanations provided by expert judges.

Initial transcript coding involved review and analysis of four transcripts chosen by the researcher as representative of all interviews conducted. The initial coding scheme and four selected interview transcripts were imported using NVivo 10 software and "nodes" were created using the initial coding scheme. The transcripts were then coded using the initial criteria, and any potential new codes and sub-codes that emerged were coded accordingly on appropriate transcript portions. Coding was also used to help to identify coherent themes, and initial codes were modified, expanded, collapsed, and refined as data collection and analysis progressed, as new themes and patterns emerged.

III.6.5 *Check Coding*

A check coding process was utilized in an effort to increase both the accuracy and trustworthiness of the analysis. The four transcripts initially selected for check coding and the accompanying coding sheet were provided to two independent researchers who had no connection to the study. The two check coders who were selected to review the transcripts were trained researchers identified and hired through a freelancer website. Both of the coders have

doctorate degrees, one in experimental psychology and the other in public health, and both have published articles in their disciplines. The two check coders separately coded the four transcripts using the preliminary codes and added new or different codes as they felt appropriate. Preliminary review indicated overall agreement among coders of between 64% and 85% for the four transcripts. Discussion between the coders revealed that most of the differences related to assign multiple codes to the same data and choices between closely related codes. As a result, some codes were collapsed into broader concepts while other codes were divided into more discrete elements. After several discussions, all coders reached general agreement on the main and subsidiary codes and coding of the data. All remaining transcripts were coded by the researcher using the agreed coding scheme and coding approach.

III.6.6 *Ethical considerations*

All participants consented to be interviewed for the study and to have their interviews audiotaped. A consent form, approved by the university Institutional Review Board, was provided to each participant before the interviews. To preserve the confidentiality of the participants, the awards program and the data provided, all interview tapes, notes from observations and interviews, and collateral data collected were kept in a locked cabinet or in a password protected file on the researcher's personal computer. When the interviews were transcribed, code names were used to identify the participants in place of their real names and, where reported, code names are used. All judging scores received from the awards program were anonymous, with numbers assigned to individual judges and all entries identified by item numbers only.

III.7 Issues of Trustworthiness

For qualitative research, issues of trustworthiness surround the credibility, dependability, confirmability, and transferability of the study results. Credibility relates to whether the research findings are accurate and credible reflections of information provided by the participants or otherwise obtained (Miles & Huberman, 1994). Credibility also relates to whether the methodology chosen is logically related to the objectives of the study. In this study, efforts were taken to improve the credibility of the research design, including obtaining data from multiple sources, soliciting feedback from industry members to narrow and refine the field of inquiry, actively searching for contrary findings and contradictory data, applying different research methods to obtain data, and reviewing the findings with colleagues to confirm the data collected.

Dependability for qualitative research involves an assurance that the findings are consistent with the data. Such dependability requires objective reviews of the data and development of procedures to reduce researcher bias and identifying data inconsistencies (Lincoln & Guba, 2000). In this study, several efforts were made to increase dependability of the findings, including the use of check coding by individuals with doctorate-level research experience who were not connected to the current study, improvement of the coding and refinement of the coding scheme based on input from independent coders, maintaining a journal of the research and decision-making processes, as well as using the data display and conclusion drawing and verification techniques recommended by Miles & Huberman (1994). Transferability of the study results is concerned with how the findings of his study may be used in other contexts. Transferability is improved when the data are shared in a rich, detailed manner that allows the reader to fully understand the context and information to determine how study findings might be usefully applied in other settings. To improve transferability of this study, verbatim quotes of

participants' responses along with detailed explanations of the context of the judging process, observations and awards program are provided.

III.8 Anticipated Limitations and Delimitations

As with any research, this study was anticipated to involve several limitations inherent in the nature of the research approach and the methods employed. First, since the study is necessarily limited to a small number (12) of recognized creative domain experts operating at a very high level of performance, it is difficult to generalize to a larger population of experts from other domains or to individuals without this recognized level of expertise. However, the 12 participants represented a broad spectrum of creative disciplines and each had significant experience in judging creativity both professionally and in awards contests, as well as decades of professional experience as a creative employee and manager. The number of participants was deemed sufficient in light of the limited number of experts in the field, the depth of information each provided about their processes and understanding of the issues, their level of expertise in the field, and the fact that multiple participants provided similar information. This similarity in participant responses indicated agreement on the main issues and relative complete coverage of the data obtained. Second, as this qualitative study involved events and actions viewed retrospectively, and using information relayed to the researcher during individual interviews of participants, it may be subject to biases or gaps in memory and perception of the participants, lack of conscious understanding of the nature of the processes employed, and filtering or efforts to appear socially acceptable or helpful to the research effort. Numerous techniques previously discussed were employed to reduce filtering and biases in memory, including semi-directed interviews, real-time think-aloud protocols and simulations. Third, this study is limited by the single researcher's ability to analyze and interpret the observed and collected data, as well as the

researcher's own cognitive biases and perceptions from the researcher's professional experience in the domain of study. Several techniques described in this section were used to maintain objectivity, including reviewing the data and conclusions with third parties in both academia and creative industries, and the use of independent data coders.

IV FINDINGS

IV.1 Introduction

The purpose of this field study was to investigate the processes and criteria domain experts use in assessing the output creativity by individuals who regularly create products in real-world settings. To facilitate this investigation, expert judges from an established international creativity awards program were observed in a field setting and participated in semi-structured interviews. Understanding how domain experts make judgments about professionals' creatively in their products and ideas, in situations where judges face significant time pressure, uncertainty, and high stakes, provides useful insights into how creativity is assessed in real-world situations. These insights may also increase understanding of the challenges that creativity researchers face in assessing creative output in experimental studies, suggest refinements to assessment methods and practices to improve the overall reliability and validity of creativity research, and improve the discernment of creativity in organizational settings.

This chapter provides the main from the interviews field observations, real-time exercises, collateral data, and the results from inter-rater reliability testing of judges' scores both from the observed contest and prior creativity award contests. Five major findings were gleaned from a detailed analysis of the research data:

1. All (100%) of the participants reported they did not specifically use the "novel and useful" definition of creativity that is prevalent in research, with most stating the definition was inadequate and incomplete for judging creativity. Instead, participants reported they look for a combination of as many as six distinct elements as key indicators of creative output: 1) uniqueness, novelty or surprise; 2) inventiveness or advancement of the state of the art; 3) overall quality of execution; 4) conceptual impact or cleverness of idea; 5) level of artisanship necessary to achieve the conceptual goal(s); and 6) utility, usefulness or appropriateness to the task.

2. A large majority (10 of 12 [83%]) of participants reported they use the same or very similar processes, criteria and standards to assess creative products and ideas in their workplaces as they do in creativity award contests.

3. The majority (9 of 12 [75%]) of participants reported that highly creative products and ideas produce an "immediate" positive or "gut" reaction, often accompanied by a strong emotional and/or physical response (sometimes termed "goose bumps") and stated the most highly creative work is readily identifiable, even among very high-level professional creative work, indicating intuition is a significant element of the participant's assessment of creativity. However, more than half of participants (7 of 12) reported also using some form of rational analysis to assess the level of creativity after their initial response, and to modify or confirm their initial intuitive assessments.

4. All of the participants (100%) stated they believed novices would find it very difficult to assess professional creative products and ideas, identified several types of mistakes the participants themselves had made as novices, and indicated they believed non-experts would be inclined to make similar types of errors (termed by some as "the novice trap") in attempting to assess professional creativity.

5. Most participants (7 of 12 [58%]) acknowledged using various heuristics or rules of thumb to judge entries, and stated they were aware of several types of biases that might affect their judgment; however, participants also believed they were able to reduce the effect of these biases through their substantial judging experience, deliberate mindfulness of the potential for biases, and the use of the Consensual Assessment Technique.

The following sections discuss each of the findings in more detail and draw support from the pilot study data and interviews of creative managers, Applied Cognitive Task Analysis

interviews of the 12 participants, direct observations by the researcher during a three-day judging event, and in-person judging simulations and think-aloud exercises with five of the participants. In total, 18 hours of recorded interviews resulted in more than 340 pages of typed transcripts that were coded and analyzed. Illustrative quotations from the interview transcripts are provided below to allow the reader a deeper view into the experiences and perspectives of the study participants in their own words; however, pseudonyms are used to protect the identities of the participants. In addition, the researcher's observations of the judging event and the responses of participants in the simulations and think-aloud protocols were recorded in contemporaneous notes, subsequently coded and analyzed, and are interwoven into the findings to provide a broader perspective of the cognitive and behavioral aspects of the activities observed. Information gleaned from program administrators was also included to provide further context of the field setting, the participants, and their behaviors. Finally, a statistical summary and the results of inter-rater reliability testing of judges' scores assessed on 63,800 items in eight awards' program contests over five years are provided.

IV.2 Findings from interviews, observations and exercises

> "The guy who invented the first wheel, he was an idiot. The guy who invented the other three, he was a genius." - Sid Caesar (1922-2014)

Finding 1: All (100%) of the participants reported they did not specifically use the "novel and useful" definition of creativity that is prevalent in research, with most stating the definition was inadequate and incomplete for judging creativity. Instead, participants reported they look for a combination of as many as six distinct elements as key indicators of creative output: 1) uniqueness, novelty or surprise; 2) inventiveness or advancement of the state of the art; 3) overall quality of execution; 4) conceptual impact or cleverness of idea; 5) level of

artisanship necessary to achieve the conceptual goal(s); and 6) utility, usefulness or appropriateness to the task.

The primary finding of this research study is that participants did not believe the "novel and useful" definition of creativity relied upon by researchers was adequate, reflective, or practicable in assessing creativity in professional award contests or in workplace settings. All of the participants felt strongly that defining creativity as being simply "novel and useful" or "unique and appropriate" failed to capture the essence, scale or totality of the concept of creativity as used in practice.

> The definition [novel and useful] is definitely narrow—it feels like textbook. It's not reality. I mean, if I was looking at a bunch of collateral brochures, let's say, what would be novel with all of those? Well, it would be one that somehow is different from the rest, right? Then you start looking for differences: "Oh, I guess this one's novel because it opens landscape versus traditional." You'd have to look for some reason to call it novel. And 'useful', well they all serve the same purpose, so I don't know if you could say that one is more useful than the other in that context. (Christine)

During interviews, most of the participants denied using any particular definition or criteria for judging creativity in contest situations. Moreover, during think-aloud exercises as the judges scored actual entries in the contest, several participants became frustrated trying to explain their process or identify the criteria being used. One participant described the difficulty in explaining his criteria or steps during one think-aloud exercise as "dumbing it down":

> Jeremy: "Ok, that one is really good."
>
> Researcher: "Why?"

Jeremy: "It just…everything works. I'm describing this to you like I'm in school again, and I'm explaining it to my professor or a teacher why I think this design is worthy. But I'm really dumbing it down…no offense…because it's the only way I know how to say it. There aren't any words I can use to express what I'm feeling and thinking right now, the combination of all these things and why it matters."

During simulated judging exercises using entries that had already been scored, several other participants also stated they could not describe the process of judging creativity or separately identify all of the criteria they used in the judging process. One participant expressed the difficulty as having "no language" to describe what she was thinking:

> Well, there is no language to attribute what you do when judging…and there is no way to separate the various component parts that make up an ultimate impression or decision in your mind. No one breaks it apart like that. You might be able to say where it fails, or that it just doesn't rise to the level, but you can't explain [why something is creative] in words. (Angela)

Other participants pointed to the significant time pressure involved in judging a large number of items, and the importance of the contest to the entrants as both contributing to an environment of high stress and insufficient time to do more than briefly review and quickly score each entry.

While all participants felt the "novel and useful" definition of creativity used in research is inadequate and incomplete for their purposes in assessing creativity in practice, through the use of Applied Cognitive Task Analysis techniques (structured probe questions, knowledge audits, and creation of lists of cognitive demands, see Appendix 7.1) a number of descriptions were elicited from participants with regard to the most important characteristics judges look for when assessing particularly creative items in a contest setting.

In an effort to better understand the criteria that experts look for in creative entries, participants were asked as part of the Applied Cognitive Task Analysis interviews to describe examples of entries that had been scored very high or that had failed to score high for one or more particular reason. While the participants' descriptions of the specific elements they recalled about memorable entries where not all encompassing or complete, and did not necessarily generalize across different contest categories, the descriptions began to uncover some of the criteria, cues, and patterns experts appeared to use when assessing creativity. As participants became more aware of the elements they could recall, they began to develop explanations for how they made their decisions.

In general terms, several participants suggested the criteria for an award winning entry is simply whether the entry is more creative than anything else the judge has seen in that category, with the top-rated entries simply perceived as "better than" what that judge might have been able to create facing the same tasks, using their own skill and experience as the standard:

> Everybody is looking for fresh. Everybody wants to be "wowed." If you want the single biggest criteria for a contest it's: "Astonish me." That's what you want; you want to say to yourself, "God, I wish I had done that! That was great!" That doesn't happen often, but it happens. (Bill)

> "I know something is really creative when I think to myself, 'I don't even know if I could have ever come up with that.'" (Leonard)

The standard of "I wish I had done that" was also the description at the top of the "judging criteria" form given to participants at the judging event. According to the director of the awards program, that quotation was incorporated into the form after she heard several judges use it in prior contests to describe the most creative entries.

When asked to recall examples of the most creative entries they had seen, most participants described two aspects of creative work that combine broad concepts of "idea" and "execution," and stated that both should be present to a significant degree in order for an item to be considered highly creative:

> Good creativity is part great design and part quality execution. A great idea with bad execution would get a 5 [out of ten], same with great execution but not a great idea. Great idea and great execution is going to get a 9 or a 10. (Angela)

Other participants referred to a description of creativity that was similar to the "novel and useful" definition used in creativity research, but with the additional elements of functionality, ingenuity, overall concept, engagement and cleverness:

> You're walking that fine line with the architecture you create that needs to be very inspired, but then it needs to be very functional. And if you look at it the functionality is key, because if it's not functional people won't fall in love with it. And then on top of that, I feel it's the creativity because if you do something that is creative, but not functional, then it is art, and I'm not judging art. It's how uniquely someone solves a problem. Or how do you sell a very common method in an uncommon and sticky manner that just left all of us speechless or it sneaked up from behind, so we are just like, "wow, this is really different!" (Fritz)

> To me a great creative work is never just one aspect, it's the whole package…the criteria I use is 'concept'…it's clever, it's smart, it's using wit or humor or intelligence or creativity in both the look, but also in the words. And then, depending on what it is, I start looking at the typography, the colors used, the photography, the illustration, how the words are placed

> on the page, and then the balance and proportion and all of those more graphic design rules start to come into play. (Christine)
>
> To me, it's the concept. I try to look at it as the consumer or user of the product, would it engage me. That's difficult sometimes because I may not be the target audience; so I have to put myself in their position, try to make it up. Is it engaging from an execution standpoint, is the layout clean, does it follow good rules of design, or does it diverge in a way that is interesting and appealing, does it make me want to pick it up and learn more? (Leonard)
>
> I think if it's too confusing or hard to "get" then it probably isn't very creative. It can't be so different and bizarre that it doesn't make sense. Especially in packaging or marketing, it has to get to the point quickly and in a clever way; it has to cause a response, otherwise it may be beautiful, but it's worthless. (Edward)

Other participants appeared to identify three components of newness, beauty, and "interesting" functionality as the key factors.

> I am looking for three things: Is it unique? Is it solving the problem in an interesting way? Is it visually appealing? (Richard)

One participant shared an example of an entry that had won top honors in a prior contest with these three components:

> An example of something I found truly creative the first time I saw it was Perlex, which is for a web page, and as you scroll down certain images appear to float up and they move and you get multiple layers happening. That's probably been five or six years out on the market now, but when I first saw that in a competition it was, "wow—amazing! This is

> totally new!" And it won because we hadn't seen it before, it was beautiful, and it did something really interesting. (Fritz)

Several other participants indicated they were more focused on the "quality" of key elements in a creative product or idea:

> In our studio, we have three things we are looking for: quality of the idea, quality of the design, and then quality of the execution. So all of those things are critical for us and everything that we judge before it goes to the client has to meet those standards. And one of the things that creates a "wow" moment is when the quality of the idea knocks you out. And then, of course, common to all is quality of the execution. If you have great ideas and brilliant design that's badly executed, they're out the door. (Bill)

> I'm looking for great concept, great execution, and great subject matter. I'm looking for consistency, messaging, layout and design, photography or illustration, does it all work, that's all part of it. If it's done technically correct, maybe it's perfect, but if falls flat on the messaging or it's boring, it doesn't communicate, then it shouldn't win an award. (David).

Some participants concentrated more on the challenge presented to the designer, and how they applied ingenuity, concept, elegance, and inspiration, to the overall execution:

> For me, the main question is: how much of a creative challenge was it? The harder the challenge, and the more the entry achieved that challenge in an inspiring, elegant and thoughtful way, the more points I'm going to give it. And then it's whether I've seen anything like it before, does it open a realm of possibilities or thinking that I've never had before. Last is the execution, when I am blown away with what they were able to achieve technically. (Jeremy)

> It has to be different, thought provoking. Everything's been done before, but something that's done in a different manner, where maybe the execution with the end product is something that is going to make me say, "Wow, I wish I had done that!" Or, "wow, how did they do that?" That's going to get my highest score. (Martin)

With regard to whether items in a contest are judged against a minimum standard, most participants indicated entries were assessed in comparison to current work in creative industries and needed to show a significant advancement in the "state of the art," in addition to being measured against all other items in the category or contest:

> Definitely there is a minimum [level of creativity] required, in a sense; to get an award, it has to be creative compared to the rest of the industry, it can't just be slightly more creative than the other items to win. You have to compare the entries to what you've seen outside of the contest to see if it's really different from other stuff out there. If none of the entries are really creative, I won't score any of them highly. (Christine)

> I may see several different entries that are interesting or unique for different reasons. I usually know right away if something is really creative, so I will remember that. But when I score the items, I don't give one item an 8 and then compare the others from there. Each entry gets a score based on the overall execution compared to what I know from the industry. (Martin)

> We all get copies of the books that are published with all of the winners from the various awards programs out there, and we are constantly looking at what others are doing in the industry, so we have a pretty good idea of what the industry is doing. (Bill)

Most of the participants agreed that while their judgments are, to a degree, based on their subjective opinions of what is visually appealing, they attempt to score the "execution" or technical achievement of contest entries using a more objective perspective of the level of artisanship execution, delivery of concept, and effort involved in the work:

> I think I make more of an objective decision first, because I know what it takes to do something, to put together a magazine cover or advertisement. You have to be able to pick out a photo, pick the right colors and the font, and those are all disciplines where if at the end its not done well and not compelling, I won't score it very high. If it's done well and is creative…it may not be my cup of tea…I don't particularly like supra ultra modern but, if there's a compelling magazine cover that leverages those types of tools and aesthetic and it's done really well, then I would score it high. (Amy)A really good technical entry will kind of become the measuring stick and so all the other entries will have to beat that to win, but that's why it's a contest. And if there are other entries that are as good or better, then we have to figure out which is the top. But if all the other entries are technically really weak then it doesn't matter. (Edward)

As discussed in Finding 4 below, during the Knowledge Audit stage of Applied Cognitive Task Analysis, some participants were able to more clearly describe and explain the criteria they use to identify creativity in professional work.

Finding 2: A large majority (10 of 12 [83%]) of participants reported they use the same or very similar processes, criteria, and standards to assess creative products and ideas in their workplaces as they do in creativity award contests.

> My approach tends to be [what] I call 'fly-by.' Let's say there's 20 items in whatever category we're viewing. I'll do a fly-by and just see if anything stands out to me. And so

> then maybe I narrow it down to five—the most attractive, or the most creative, and then I'll come back and then take a closer look at each of those five. And then whittle it down from there. So, it's an elimination process, I guess. But that's pretty much how I do it in our agency too, just on a smaller scale. (Christine)
>
> For me, personally, judging is no different. It's like when I meet my team as creative director, and I see seven different ideas, and it's going to take me three minutes to just go through them and say, "no, no, no, yes, and here's why, and let's move forward with this and why don't you take it for a spin on that idea?" (Fritz)

Some of the differences between judging awards contests and assessing creativity in the workplace the participants revealed centered around the amount of time available to review each item, the availability of background information on the client or problem presented, and the ability to discuss the output during the creative process in the workplace which is not possible in a contest setting:

> The process is different because there is probably more communication back and forth in an agency setting. If somebody in the office says: "which of these do you like?" I'm probably going to be inquisitive about it, where in an awards show you don't have that ability. The other thing with the show is, you're maybe in the first section, you're judging, and you're looking down the room and you've got 50 more sections you have to get through by 3 p.m. or something, and so you're like: Ok, I have to set time limits for myself that I have to keep moving. (Christine)
>
> Interviewer: How is judging entries in creativity competitions different than how you decide whether something is creative in your agency?

> Leonard: I wouldn't say it's different at all. I'm just critical. The only difference is maybe I take more time, obviously, internally with my team to make sure that they are following the same rules and make changes.
>
> Sometimes the entry will have a short paragraph on the back or another page written by the designer or team that explains a little of the background, what they were trying to accomplish, maybe why they chose a certain approach. I like to read those because I get a better understanding of the challenge that was presented and how the contestant looked at it. That kind of information makes it more like my role at my agency where I'm usually critiquing the work of others with a limited amount of background on the client and project brief. (Edward)

The researcher also observed the process of some participants during the three-day judging event. The 2014 judging event was conducted in a hotel ballroom in a major metropolitan city on the East Coast of the United States. Eight judges had been selected for the event by the program director, and each traveled to the venue at their own expense. The awards program paid for the judges' accommodations and provided meals during the event; however, none of the judges were compensated for their participation. During a round-table discussion after the event, many of the judges said they participated in judging events out of a sense of obligation to the industry, as a way of staying connected to the very best work in their profession, and to strengthen their professional credentials.

At the beginning of the judging event, the judges gathered to receive instructions and scoring devices in large ballroom where entries were laid out on long tables separated by category. Each judge was given an electronic tablet and a one-page explanation of how to use it in the judging process. After the program director generally described the judging criteria on the forms, each

judge was shown how to input their scores on the tablet device, and the group was told they had three days to score all of the entries (in excess of 800 entries in more than 60 categories), and to separately mark on the tablet any entries they wished to nominate for Best in Show. The judges received no other instructions or directions on how to judge the entries, how to go about reviewing the categories, or what was expected of them.

Once released to start reviewing items, judges moved slowly around the ballroom, appearing to drift aimlessly among the entries and tables. About 15 minutes after moving among the tables, one judge sat down at a table and began perusing a group of entries held in a binder, but she did not appear to be scoring any of the items. A few minutes later, another judge sat down at a different table and he too began thumbing through the items but did not appear to be entering any scores on his tablet. Other judges appeared to look at only a few items at a table before moving on to another. After about 25 minutes, the first of the eight judges appeared to start entering scores on his tablet device. Within 45 minutes, all the judges were reviewing items and entering scores. However, during the first hour of the judging event, not one of the judges was seen or heard speaking to any other judge. One of the participants later explained why he took this approach:

> In general, I'm trying to [judge] alone…I've found for some [of the categories] you start seeing herds of people. But I try to avoid that because you do get swayed, as much as you try not. You'll hear people saying: "Oh, this is amazing, did you see this!?" and I don't want to hear it; it does play into your psyche…And that is the risk of doing it as a group. (Leonard)

Once the scoring process began in earnest, a pattern seemed to emerge among the judges. Judges moved around until they came to a category that was physically away from the other

judges and then either stand or sit at the table where the entries were placed. Several judges were observed looking almost casually at some of the items in the category, handling and viewing each for only a few seconds before moving to the next. After several or all of the items were viewed in this way (usually depending on the number of entries in the category), the judges would review the items again and started to enter scores on the tablet. When all the scores were entered, the judges moved away from the table, and the process would repeat wherein judges moved around each other politely but maintained their distance. One of the participants explained why he preferred this approach:

> I am a packaging and graphic designer, I prefer to see the physical version of the entries over seeing them online or just a photo. That's why I like this awards program, all of the entries are physically here and I can touch them and compare them. And I can review entries at my pace without being influenced in any way by other judges. (Richard)

The first time any of the judges talked to each after the event began was outside the ballroom during a spontaneous break. However, during the first day, the judges appeared to make only small talk during breaks and over lunch; there was no discussion of the contest entries. In fact, the first time the researcher overheard any judge discussing an entry was in the afternoon on the second day of judging. In this instance, one of the judges had just walked away from a category after scoring and was approached by another judge who asked, "Are you all done with that category?" When the other judge replied yes, the two judges discussed one entry in that category both had already scored and agreed they thought it should be considered for Best in Show. As the judging process continued that second day, more judges were overheard discussing specific entries or categories, usually in the context of a particular entry that might be considered for a top award. This judging and scoring process, primarily conducted in isolation,

with breaks to discuss entries that they had already scored, continued over the contest's three days.

During breaks in judging, the researcher was able to discuss the contest with all of the judges, and five of the judges consented to participate in the study and be interviewed. Four of the participants also agreed to take part in think-aloud exercises while judging contest entries in one of the categories or during simulated judging exercises using entries they had already scored. After the exercises concluded, these four participants were interviewed and provided more specifics about the process used in judging contest entries. Several participants confirmed that it is important that they judge separately to reduce the possibility of being influenced, and that they try to assess the level of creativity in the contest before scoring any items:

> So I look over all of the entries in a particular category first, as I said, then I go back and do the scoring. I want to get an idea of the overall quality and creativity, and if I see something really great, I will remember it, and then score it more highly when I come back to do the actual numbers. (James)

> The contest gives us a rating sheet and, basically, a 1 or 2 has no creativity at all, a 9-10 is at the very top, it is the best of the category and maybe one of the best in the whole show. If something is good, but not quite as good as it could be, maybe the coloring wasn't quite right or the concept was a little off or just another version of something else we've seen before, then it will still get a 5 or 6, but I can't give it a top score. (Martin)

During applied cognitive task analysis dialogues with participants about the cognitive effort required to judge creativity, many identified the most challenging part to be separating and ranking the very best entries from others that are very good. Most participants stated they felt different entries may be highly creative for different reasons and comparing them was almost

impossible, especially if for different categories or if they have different levels of complexity. Below a score of 7, all of the participants said the decision was rather straightforward, and that they tend to gravitate toward giving scores of around 3 to 5 at the low end to make a clear separation between those worthy of consideration and those below the "highly creative" threshold.

In fact, not all of the judges used the entire 1 to 10 scale suggested by the contest administrators (see summary statistical analysis of judges' scores in Table 1 of Section 4.3 below). One of the participants in this study stated he rarely gave an entry a score of 10, as he felt the top score should be reserved for only the very best of the best, sharing that he uses the following quote as his guide for score of 10: "'Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away.' [Antoine de Saint-Exupery (1900 - 1944)] The longer you've been in the business, the fewer things get a 10 because you've seen a lot." (Edward). Another participant, Amy, said she rarely gives scores below a 5 because she knows all entries were created by professionals so they must be at least "average."

While observing the event, the researcher noticed variation in the amount of time judges spent on viewing and scoring each item. To test this variation, on the second day when the judges appeared to have become proficient at using the tablet for scoring, the researcher attempted to surreptitiously record the amount of time four different judges each spent looking at individual entries in various categories and noted how long it took those judges to enter their scores. Depending on the category, and allowing for the inexact nature of timing individuals from a distance, all four judges appeared to spend less than two seconds reviewing each item that consisted of a single page in a binder or of a single item on a table. A similar amount of time (two seconds) was generally spent recording scores, but in a few instances, judges were observed

viewing up to six entries in a row and then entering scores for all six items at the same time. Sometimes, judges spent considerably more time on single item, usually between five and eight seconds, and in some cases more than 10 seconds, looking at the entry more thoroughly and critically. Interestingly, none of the judges were timed spending between two and five seconds to evaluate any of the entries. When asked individually after the scoring was complete about the instances when they took more time, judges indicated those particular items appeared more creative and potentially worthy of a higher score, and the judges wanted to spend more time considering the entry to be sure, in particular looking at execution and detail. One of the participants explained his thinking:

> Usually, you can tell really quickly if something is good or great, it stands out. So then I'm just trying to decide whether it gets an 8 or 9 or whatever. The other entries that aren't very good, there's no need to spend any time on those. But some categories have more going into them, like big corporate reports or a video, you have to go through those and see all of it, so those categories do take longer. (Richard)

Only one participant (Jeremy) said he would sometimes go back to other items after he began judging and modify his previous scores based on the quality of subsequent entries viewed. Most of the other participants said they didn't have time to go back and revise their scores, and all participants said they felt very confident that the scores they awarded were correct based on their experience and knowledge of the industry and in comparison with other items in the contest.

> I'm not going to lie, first impressions matter. If something doesn't get my attention right away with the quality of work and ideas presented, I'm probably not going to spend the time to dig through it to find something creative about it. (Edward)

A few participants, however, indicated the standard they used might change depending on the level of creativity being shown at the contest.

> I feel that at the beginning of a show obviously the first pieces are much more about my professional viewpoint. And then as I go let's say half a day through an award show… I'm already going to be jaded by my background of that half day. So if I see really bad work that half day, suddenly something I thought was a five might suddenly turn into a seven just because I've seen so much bad stuff [laughter] but in the end it's all relative to the other entries. (Fritz)

> If it's really fantastic, even with crappy presentation, I'll give it an award, if it's that amazing and cutting edge. So I'm not saying that I'm never going to give anything an award if it's sloppily done, but me, personally, I have a hard time looking past that, because it's a given. If you're going to take the time to enter something in a contest, then surely you can do a credible, decent job with it. (David).

Finding 3: The majority (9 of 12 [75%]) of participants reported that highly creative products and ideas produce an "immediate" positive or "gut" reaction, often accompanied by a strong emotional and/or physical response (sometimes termed "goose bumps"), and stated that the most highly creative work is easily identifiable, even among professional creative work, indicating intuition is a significant element of an expert's assessment of creativity. However, more than half of participants (7 of 12) reported also using some form of rational analysis to assess the level of creativity and to modify or confirm their initial intuitive assessments.

> David: "If it hits me in the gut, it's a winner."

> It's mainly really a gut instinct. But gut instinct based on my experience, based on my knowledge, based on looking at creative work every day and being in between my own creative team and their creative instinct and egos. (Fritz)

During a think-aloud exercise while judging an actual contest category, one participant had a very sudden emotional and visible physical reaction as he turned the pages of a binder of entries and came upon one particularly creative entry and stopped:

> So, I just look at each item, kind of going through them like this and saying pretty quickly, "no…not great…that's just ok…maybe a 5." Uh…[2-3 seconds pass, the participant does not move while staring at the entry, he is visibly moved]…oh my. I, uh…wow! I mean, there…[he slaps the page of the binder]…look! [He points at his forearm] See, I have goose bumps!!" (Martin)

Most of the participants reported highly creative work often produces a strong visceral response. Amy said, "If something is really creative, I want to eat it. I want to lay in it, I want to put that all over my body and feel happy all the time." Angela concurred, saying "If it's given me delight in some way then I'm going to give it a higher rating."

Most of the participants also said their response to an item's creativity, or lack there of, was almost instantaneous, and that it was the magnitude of impact on the viewer that signaled the level of creativity:

> I have to see the effort involved…there has to be a spark of something non-traditional and an effort to express something, you know, compelling. I can tell immediately…whether it's good or bad, you know, that there was an approach that pushed it out of the ordinary, I think that is what creative means to me. (Amy)

> There are always those cases where there's something that is just like "whoa, that's amazing!" And for some reason it resonates with you. If it's purely a creativity criteria, I think usually at this point it's something that happens pretty quickly, that one's going to stand out immediately. Let's assume it's something simple, it's one page, I'm looking at each for probably no more than three seconds initially…which is, in the real world, it's two point five seconds [laughs]. I know they've put a lot of effort into putting together their submissions in, so I don't want to make a rash judgment, I want to give each entry it's due time, but usually I know it when I see it. (Christine)

> Does it work for the target audience? Does it fit the purpose of the creative brief and even though you don't get a creative feel for every single piece you look at, you have an idea if it is right for the client and consumer. So you assume a lot of things but this and that, target audience, purpose, my experience and my knowledge are mixed up and then you give a very quick instant vote. (Fritz)

However, some participants (David, Leonard, and Martin) said the speed of their process and judgments might have had more to do with the limited amount of time available to judge a very large number of entries in a contest than with the ease of identifying creative work, and that they had to accept their initial impressions as valid because they were unable to spend as much time as they might like on each entry. Others (Fritz, David, and Edward) mentioned they race through entries looking for only the very best and grade those that make an immediate impression as more creative because, "I don't want to be looking at bad work all day" (David).

While almost all of the participants agreed they often had an immediate response to highly creative work, most added they felt the immediate response needed to be tempered with a more analytical review before giving a final score:

> If it's really complex, it might slow things down, but for the most part, you're making an instantaneous judgment that's the combination of all of that experience, discipline, knowledge that you have, that creates that instant physical response, and the only word that comes out of your mouth is "wow." And then you start to take it apart. (Bill)

> Sometimes when you're looking at a piece of work on a table, I notice the executional qualities first. Call it a five second look—those kind of things. Then I start to look more closely. The concept comes first and then you get to make your way to execution and then you hope the execution actually makes the idea better than worse. Nothing depresses me more than a great idea ruined by poor execution. (James)

> There's a kind of visceral side of judging and then there's a cognitive side. From a visceral side, you see something and the question is does it make you say, "Wow!" Or does it give you an emotional reaction, does it raise the hairs on the back of your neck? But then you have the logical reaction to it, does it work like it's supposed to or what am I getting for my money? (Jeremy)

Many participants spent the additional time and analysis to consider the artisanship and execution of the concept and creative idea, and they ultimately compared the creativity with other entries:

> If it wows me or if there's a really brilliant idea there or something you know has a really unique way of showing me something that's in it. Or it surprises me, or delights me, or intrigues me, or it pulls me in. That's what just stops you and makes you take notice. But then as a judge it's your job to really take a look at the craft. You have to dissect it. I start with the idea behind the work, then I look at, "Okay, how well did they execute that idea?" (Angela)

> You can tell pretty quickly…I break it down into the nanoseconds. I think it is the visual quality first that I recognize, the visual design, the taste and finesse of the execution. The first thing that stops me is, "has it been done well?" And that's the reverse process for work at the office. It's concept first and then execution at my agency, but here where everything has been fully executed, you notice execution first and then look to see what the idea was, but it still happens very quickly. (James)
>
> Interviewer: How quickly do you recognize when an entry is very creative?
>
> Leonard: Like one second. But then I have to say, "Are my eyes deceiving me?" And then I have to look at it more closely, and I have to see everything else in that category to be sure. There may be something else that's even better.

Finding 4: All of the participants (100%) stated that they felt novices would find it very difficult to assess professional creative products and ideas, identifying several mistakes the participants themselves had made as novices, and that they believed non-experts would be inclined to make similar types of errors (termed by some "the novice trap") in attempting to assess professional creativity.

Although not a particularly surprising finding considering their level of experience and backgrounds, all of the participants felt strongly that judging creativity for both awards and for professional work requires not only extensive experience in a creative industry assessing products and ideas, but also significant skill and direct experience in doing actual creative work. For example, Amy said, "You have to some years of experience to be able to appreciate the craftsmanship enough to be objective."

> To judge creativity someone needs to be a trained designer and have creative work experience, otherwise it's just subjective opinion with no basis in the industry. They might

> just be looking for what's pretty or something they haven't seen before but everyone in the trade already has. They're not looking for technical design elements, just what they like. (Leonard)

> There are lots of other contests out there where the judges are celebrities or executives in an industry with little or no real experience, and they seem to give awards to the things that personally excite them. But unless they really know what's going on in the industry and the level of effort required to pull off a truly creative idea, they just fall for the first thing that grabs their attention or looks cool in the moment. We see it in our agency all the time: clients and junior designers will get all excited about the next "new thing," and they can't help but think it's the best. But that's the novice's trap, they can't get out of that way of thinking until they have actually been working in the field a while and considered all the other great work that's out there. (Bill)

As part of the second stage of the cognitive task analysis process, the Knowledge Audit (App. 7.1), some participants were asked why certain parts of the task might pose a challenge to novices and were encouraged to imagine the kinds of mistakes a novice or non-expert might make when judging the creative work of others. All of the participants who responded to these task analysis probes where able to draw on past experiences and conceive of multiple ways someone without the requisite level of experience might incorrectly judge a product or idea as creative, or fail to identify a highly creative entry. Interestingly, during this phase of the CTA process, participants found their voice in explaining the key criteria they look for when assessing creativity. Most of the participants' examples of novice mistakes directly related to a lack of knowledge and inability to recognize several key elements of creativity: what has come before (lack of novelty); the subtle but powerful effect a well-developed idea (conceptual impact or

cleverness); the utility or appropriateness of the item to the task presented; the level of inventiveness or advance in the state of the art represented (inventive step); the level of experience and creative skills required to accomplish the concept as reflected in the finished product (artisanship); or the delivery and impact of an entry that elevates it above the others during assessment (execution). Of particular note was the participants' belief that creativity requires a confluence of many elements, some of which might be missed or ignored by a novice:

> I think someone without the right level of experience would miss the subtle aspects of something really creative, probably stopping at the point of execution, without digging deeper into the idea. Judges who don't have a lot of experience won't understand the amount of effort and difficulty required to make something new and conceptually interesting, they might think what they are seeing was easy to do but someone with a lot of experience in a creative field would know how hard that would be to pull off. (James).

A majority of the participants explained how experience in a creative field alone is insufficient to attain the level of expertise necessary to effectively judge creative products and ideas—that assessing creativity requires extensive experience in evaluating and analyzing the work of others during the creative development process. Several participants also questioned whether someone experienced in only one aspect of creative design could effectively judge the creativity of a completely different type of product or area of specialty, e.g., whether technical expertise in website design could be effectively transferred in judging creative photography. The participants who raised these concerns said a true expert in a creative field must have a combination of extensive experience working as a creative and many years deliberately practicing, getting feedback, and learning how to develop and evaluate creativity in different forms to be an effective judge of creativity:

> I think judges who don't have much experience haven't seen as much as a seasoned professional, and so they're pretty impressionable. They don't have the knowledge maybe, and the skill sets on how a lot of things are done, and they get a little too excited just because, to a novice, everything is new and fresh and cool. But just being in the business for 30 years isn't enough; if they haven't mastered the craft they shouldn't be judges, because they don't take it as seriously as a master. (David)
>
> For me, judgment is a trained muscle, and it comes from knowledge, experience and discipline. It's all of those things, it's literally a trained muscle that allows me to recognize instantly whether what I have in front of me meets all the criteria, and that only occurs over time. (Bill)
>
> It's not enough to know how to create something, you have to learn how to critique design…how to tear it down and know what's missing, what can be changed and added, to make it better. If it's as good as it can be, then you can say it's truly creative. (Angela)
>
> It takes a lot longer, a lot more work for a junior person to see the difference between something that is pretty good and something that's really good. It is just years of disciplined thinking and practice…and getting feedback from others in order to get the perspective to know what's great and what isn't. (Jeremy)

Several participants also commented on how consistent past groups of expert judges had been at identifying the most creative work in prior contests where the participants had been on the judging panel, even when that panel had not judged as a group before and each originated from vastly different creative fields, countries, and cultures:

> If you get ten judges together and every one of them is an industry professional they [tend to be very consistent]; but yet they all have very different backgrounds as far as their own

creative insights or their own intuition or why they decide something is visually pleasing or not. [Even though] it's a personal decision, I feel if you have enough people coming together with the right background… that you will get a fairly fair judgment out of it. (Fritz)

Most people who are at a level where they're going to be picked to judge an awards show, I mean they're—in theory—at the very top of their game, and so you've got a whole bunch of people who are the very best, very Type A. But the outcome is 99 times out of a 100 still going to be very consistent for the top awards. (James)

Everything is subjective, obviously, and if you get a room full of people together they can't even figure out what to get on a pizza. But if you have eight people who are strong personalities and great in their respective fields at what they do, if the work is that good and in this contest it is, the cream always rises. (Leonard)

If you can't stop thinking about an entry or can't wait to talk to other people about it, that's one I would give the highest score to. I think it's tricky because the volume of entries is daunting, and you start to get tired…things start to meld together. But there are usually four or five overall that are just clearly the best. And at the end of the show, all the judges got together, and it was amazing how much we all agreed that those four or five items were among the best. We had totally different backgrounds and different personal aesthetics, with people from different countries and different cultures on the panel, and yet we gravitated towards the same items without even knowing it. (Amy)

There's an old expression that is: a judge's opinion can change by what he ate for breakfast. And so it is totally subjective, and I've often wondered if I went back a week later or a month later or a even a year later with the exact same work, if I would give it the

> exact same marks. And in most cases, if the work is that fantastic, a top award winner, I think the yes, I would. (David).

As will be discussed in Chapter 4.3 below, the judges' actual scores for this contest, and for seven prior contests, were analyzed for inter-rater reliability to test the judges' beliefs about the consistency and reliability of group scoring. While the judges' scores for those eight different contests showed a high level of inter-rater reliability, their agreement did not reach the "very high" level some participants might have assumed. The possible reasons for this are discussed in Section 5 below.

Finding 5: Most participants (7 of 12 [58%]) acknowledged using various heuristics or rules of thumb to judge entries and were aware of several types of biases that might affect their judgment; however, participants also believed they were able to reduce the effect of most such biases through their substantial judging experience, deliberate mindfulness of the potential for biases, and the use of the Consensual Assessment Technique.

> I think the honest answer is: I don't think you can prevent [biases from] affecting your choices. I think we walk around with those in our head all the time. I want to have an open mind…but it's near impossible. You have to be aware of those biases and work through that and try to be fair with what you're looking at. (Christine)

As noted above, all judges appeared to deliberately work alone during the event, only discussing entries after they had been scored. Many of the participants indicated this was part of their strategy to avoid being "swayed" by other judges' reactions, which they felt might bias them to award a higher or lower score. However, many participants also acknowledged that they tended to review several items in different categories before starting the actual scoring process,

in order to get a sense of the overall level of creativity in the contest, despite the possibility of pre-judging entries based on a small initial sample:

> At the beginning of the judging process, I'll skip ahead to get a sense of what's to come, but then once I've gone through a few categories, I do the rest of them in order. I get a sense of the kind of submissions the show is getting, and then I start judging the entries. (James)

> I do a quick round-up of the room. Just to see where I want to start and what type of work is there. I'm not being critical at all at that point. Obviously, certain things catch my eye, and I'm like, "Wow, that's really cool, I can't wait to get to that one. But I try to withhold judgment and see more of what's there." (Leonard)

All the participants expressed awareness of the potential for various biases to impact their assessment of creativity both in awards contests and in the workplace. However, they agreed that being mindful of the potential for bias and working hard to assess creativity objectively, using their past experiences and expertise as guides, reduced the possibility of bias impacting their ability to identify the most creative products and ideas and score them accordingly. Importantly, all of the participants also pointed to the fact that they were judging entries individually, and that their scores would be combined with the scores of other experts as reassurance that any bias in individual judgments would be effectively rectified in the final contest score averages.

Most participants also felt the types of biases and heuristics inherent in the judging process would be more significant, and perhaps insurmountable, challenges for non-experts, particularly in situations of significant time pressure, high stakes, and uncertainty such as an awards contest:

> One of the things you see [with inexperienced judges] is everyone tends to find something they fall in love with early on and … everything else systematically is graded lower or higher against that one. I was guilty of that at first where I would start on print ads or magazine covers and be like, that's really cool, and it's the first thing I saw that was good so everything else measured against that. To prevent that from happening, I have to remember how it may have affected me in past situations, just be mindful of the impact those thoughts can have, and go through the steps that I know will help reduce that kind of thinking from affecting my decisions. (Amy)
>
> These entries are from real people with real jobs and whether they win or lose this contest can have a major impact on their career. I have to be very careful to not to get in a rush or let the work of other entries affect my judgment; each entry has to be considered on its own merit before it is compared to the others. I keep that in mind as I go through each one, almost methodically, even if I can only spend ten seconds looking at something, so I don't get caught up in being too critical or not critical enough. (Jeremy)

Lastly, participants reported that the use of multiple expert judges, who evaluated all items independently and based on their extensive judging and industry experience, helped to reduce the likelihood of impact of pure subjectivity and process biases. Participants embraced and approved the Consensual Assessment Technique as an appropriate method to judge creativity and to establish high validity and reliability of results in both contests and professional practice.

IV.3 Judges' scoring and interrater agreement testing

As described in the methodology chapter, scoring data was obtained from the creativity awards program that included all entries for the contest of this study and 9 other contests from 2009-2014. In each of those years, the program held separate contests for digital and interactive

media at one part of the year, and print media and packaging, etc., at another point, each with different entries and different judging panels, resulting in two sets of judges' scores per year. The data was received in archival form, with the names of the judges made anonymous with random ID codes, and individual judge's scores of 0 to 10 assigned to each entry. Due to apparent errors in the random assignment of some entry numbers, however, only scores from eight contests, including the contest involved in this study, were suitable for analysis. For the eight contests included in the analysis, 70 judges scored a total of 8,699 entries. Approval to use the archival data for research purposes was separately obtained from the university's Institutional Review Board.

The purpose of obtaining this scoring data was to test the inter-rater agreement of the judges over several years. However, to protect judges' confidentiality, data for all contests over a five-year period were provided in an anonymous format, and since the participants had acted as judges in different contests over that five-year period, it was not possible to specifically test the level of agreement of any single judge or group of judges with other judges in a particular contest or to compare all of the participants' scores across contests. Nonetheless, the data provides interesting insights into the level of agreement expert judges achieve in actual judging environments over several successive years, in contests where the participants acted as judges at different points, and allows the results from this field setting to be compared to the results of published experimental creativity research studies.

In creativity research, Intra-Class Correlation (ICC) analysis is a primary method to measure inter-rater agreement among judges. ICC may be performed under different models depending on the raters and the items rated. The model of ICC analysis depends first on whether the judges form the population of all judges of interest or if they are taken as a random sample of

all possible judges. Second, the model of ICC depends on whether all subjects or items rated form the target population or are based on a random sample. Lastly, the model of ICC depends on the whether the reliability is based on individual ratings or mean ratings of all judges (Shrout and Fleiss, 1979). These considerations give rise to three different forms of models on which ICC is based.

One-way random effects model. This model is most appropriate when judges are taken as a random sample from a population of possible judges, who rate all subjects of interest. In this model, judges are treated as a random sample, and the focus of interest is a one-factor ANOVA test to determine whether there is a significant subject effect. This model applies even when the researcher cannot associate a particular subject with a particular rater because information is lacking about which judge assigned which score to a subject.

Two-way random effects model. In this model, both judges and subjects are comprised of random samples from respective populations of judges and subjects. Judges rate all (N) subjects chosen at random from a population of subjects, and it is known how each judge rated each subject. In this model, the ICC is interpreted as the proportion of subject added to the judge variance that is associated with differences among the scores of the subjects. The ICC is interpreted as generalizable to all possible judges.

Two-way mixed model. In this model, all judges of a population rate a random-sample of subjects from a well-defined population. This particular model is a mixed model as judges are treated as fixed effect (not as a random sample of all possible judges) and the targets treated as random effect in the model. In this model, the ICC coefficient is equivalent to the two-way random effects model, but the only difference is that ICC computed and tested is not generalizable beyond the given set of judges.

For each of the three models, the type of ICC computation method also requires a choice from two alternatives: (i) whether ICC is to be computed using absolute agreement or (ii) ICC is to be computed using a consistency approach. Absolute agreement is a measure of whether judges assign the same absolute score. Absolute agreement is often used when systematic variability due to raters is relevant. In contrast, the consistency method of ICC computation measures if ratings are highly correlated, even if they are not identical in absolute terms. Consistency agreement is often used when systematic variability due to raters is irrelevant. These alternatives use different versions of the intraclass correlation coefficient—(i) Single measure reliability where individual ratings constitute the unit of analysis and (ii) Average measure reliability where the mean of all ratings is the unit of analysis. Average measure reliability is most appropriate when the research design involves averaging multiple ratings for each item and using an individual rating would involve too much uncertainty.

For this research study, a two-way random effects Intraclass Correlation model using the consistency alternative (ICC(2)) was employed to assess the inter-rater agreement in all contests because both judges and items scored were random samples, and the judges' scores are averaged across items to achieve a creativity score for the contest. The 0 to 10 ratings that judges used were assumed to be interval-level data, and the test statistic follows F distribution. A test of the significance of the ICC was also performed. The statistical hypothesis formulated to test the significance of the ICC coefficient is:

Null hypothesis H_0: ICC coefficient is not significant (= 0).

Alternate hypothesis H_1: ICC coefficient is significant (# 0).

The test for this study is performed using a .05 level of significance, thus the null hypothesis can be rejected, and the coefficient is considered significant if the p-value of the test is less .05.

Tables 3 to 11 report descriptive statistics for contests 1 to 8 respectively. The number (N) of items judged, as well as the minimum score, maximum score, mean, and standard deviation of scores are reported for each for each judge (identified separately by an anonymous number[2]) for each contest.

Table 4 Descriptive Statistics for Scores of Items in Contest 1

Judge	N	Minimum	Maximum	Mean	Std. Deviation
2	2652	0	9	5.72	1.498
3	2648	0	9	5.23	1.853
4	2154	0	9	6.11	2.023
5	2233	0	9	6.17	1.520
6	205	1	8	5.18	1.355
8	1828	0	9	7.10	1.828
9	2034	0	10	4.10	2.503
10	402	5	9	7.97	.869

[2] Judge numbers were assigned in the data set by the awards program before receipt by the researcher and were not in strict numerical order.

Table 5 Descriptive Statistics for Scores of Items in Contest 2

Judge	N	Minimum	Maximum	Mean	Std. Deviation
11	511	5	9	7.96	.884
12	492	0	9	5.84	1.837
13	511	0	9	4.98	2.380
14	509	1	9	6.62	1.730
15	504	1	9	6.50	1.897
16	512	0	9	5.58	2.677
23	651	0	9	4.73	2.036
24	487	0	9	4.72	2.122
25	501	0	9	5.16	2.616
26	411	1	9	5.63	1.767

Table 6 Descriptive Statistics for Scores of Items in Contest 3

Judge	N	Minimum	Maximum	Mean	Std. Deviation
28	511	1	9	4.54	2.080
29	558	1	9	6.01	1.443
31	527	1	10	4.99	2.006
32	511	1	9	4.42	2.105
33	538	2	9	5.57	1.610
34	530	1	9	6.41	1.356
39	636	1	10	5.95	1.526
40	583	1	9	5.15	1.727
41	649	1	9	4.90	2.199

Table 7 Descriptive Statistics for Scores of Items in Contest 4

Judge	N	Minimum	Maximum	Mean	Std. Deviation
46	1336	1	9	5.21	1.766
47	1321	1	10	4.43	2.215
51	1269	1	10	5.06	2.363
52	1326	1	10	5.57	1.662
53	1331	2	10	6.32	1.550
55	1332	1	10	4.97	2.254
56	1305	1	10	5.18	1.848
57	1334	3	10	7.62	1.415

Table 8 Descriptive Statistics for Scores of Items in Contest 5

Judge	N	Minimum	Maximum	Mean	Std. Deviation
70	978	1	9	5.56	1.296
71	966	1	10	4.84	1.732
72	982	1	10	3.62	2.258
73	983	0	10	4.75	2.250
74	937	3	10	6.15	1.243
75	981	1	10	5.12	2.483
76	977	1	10	5.80	2.222
77	970	1	10	5.44	1.892
78	976	1	10	6.31	1.530
79	915	2	10	5.68	2.301

Table 9 Descriptive Statistics for Scores of Items in Contest 6

Judge	N	Minimum	Maximum	Mean	Std. Deviation
76	465	1	9	6.44	1.571
80	487	1	10	5.15	2.103
81	487	1	10	4.93	1.782
82	381	1	10	7.16	1.512
83	487	1	10	7.15	1.973
85	487	1	9	5.85	1.808
86	487	2	10	7.07	1.972
87	481	1	10	6.85	1.478

Table 10 Descriptive Statistics for Scores of Items in Contest 7

Judge	N	Minimum	Maximum	Mean	Std. Deviation
29	913	1	10	5.89	2.671
88	956	1	10	5.03	2.700
90	964	1	10	5.61	2.886
91	977	1	10	5.21	1.688
92	951	0	10	4.60	1.540
93	969	2	10	4.36	1.658
94	927	1	10	4.91	1.857
95	963	1	10	4.90	1.603
96	955	1	10	5.82	1.413
97	971	1	10	5.36	3.140

Table 11 Descriptive Statistics for Scores of Items in Contest 8

Judge	N	Minimum	Maximum	Mean	Std. Deviation
105	859	1	10	6.61	1.687
106	854	1	10	8.26	1.258
107	863	1	10	6.02	2.474
108	861	1	10	5.30	1.931
109	835	1	10	5.39	1.516
110	858	1	9	5.96	1.583
112	864	1	10	4.60	1.951

The inter-rater agreement analysis was performed using the IBM SPSS version 19.0 software application applying a two-way random effects model (ICC(2)) consistent with recommendations of Shrout & Fleiss (1979). Listwise deletion of missing values was employed as the default in SPSS to handle instances of missing data.

The summary of the ICC analysis is presented in Table 12. In this table, summary of the computed ICC coefficient along with the test for its significance is reported for both single and average measure types of computation of ICC. In addition, 95% confidence intervals for the ICC coefficient along with the F statistic and the associated p value is reported.

Table 12 Intraclass Correlation Coefficient

Contest		Intraclass Correlation	95% Confidence Interval		F Test (True Value 0)	
			Lower Bound	Upper Bound	Value	P value
Contest 1	Single Measures	.256	.177	.333	4.036	<.001
	Average Measures	.674	.563	.750	4.036	<.001
Contest 2	Single Measures	.360	.284	.430	6.371	<.001
	Average Measures	.797	.735	.841	6.371	<.001
Contest 3	Single Measures	.294	.247	.341	3.390	<.001
	Average Measures	.676	.621	.721	3.390	<.001
Contest 4	Single Measures	.260	.201	.318	5.418	<.001
	Average Measures	.759	.694	.808	5.418	<.001
Contest 5	Single Measures	.271	.229	.313	5.559	<.001
	Average Measures	.788	.748	.820	5.559	<.001
Contest 6	Single Measures	.277	.206	.350	5.333	<.001
	Average Measures	.754	.675	.811	5.333	<.001
Contest 7	Single Measures	.188	.165	.214	3.462	<.001
	Average Measures	.699	.665	.731	3.462	<.001
Contest 8	Single Measures	.250	.162	.336	4.880	<.001
	Average Measures	.700	.574	.780	4.880	<.001

For each of the contests, the F test for the significance of the ICC coefficient reports $p < .05$, indicating a statistically significant positive inter-rater agreement among judges. Average Measures report the Intraclass Correlation for each contest, with the lowest ICC coefficient at 0.674 for contest 1 and the highest ICC at 0.797 for contest 2. Reliability coefficients of .65 to .80 are considered to indicate moderate to high agreement in creativity research studies (Amabile, 1996; Kaufman & Baer, 2012). The ICC coefficients reported translate to an effect size measure of at least 0.449 ($\eta^2 > 0.449$). This means the effect size for inter-rater agreement measure is at least moderate, a further indication of a high (but not very high) degree of consistency and agreement among judges across items for each of the contests.

V DISCUSSION

V.1 Introduction

The purpose of this multi-case field study was to investigate the criteria and processes used by expert judges in assessing the creativity of professional entries submitted to a creativity award contest. The study assumes that understanding how domain experts make judgments about creativity of professional products at awards contests, in situations where judges face significant time pressure, uncertainty, and ill-defined goals, will provide useful insights into how creativity is assessed in real-world situations. To conduct this investigation, the study employed naturalistic inquiry techniques to collect qualitative data through in-depth interviews, observations, think-aloud and simulation exercises, and used quantitative analysis to test scoring data. Participants in this study were recognized subject-matter experts in creative industries with substantial experience judging creativity in both workplace and professional awards contests. Observation of real-time entry judging and simulation exercises occurred during an actual awards contest event. Archival data of scores given in the contest observed and in eight prior creativity award contests over the previous five years were obtained after all awards were announced. The data qualitative was collected, coded, and analyzed using the conceptual frameworks identified by a literature review of relevant topics and then reorganized using themes and concepts that emerged from the data that responded to the research questions.

To explore how domain experts judge the creativity of entries in a professional awards program, this research investigated the following research questions:

1. What criteria do expert judges of creativity awards contests report using to assess the creativity of products and ideas; in particular, do experts use the “novel and useful” definition of creativity that is employed in most creativity research or some other criteria?

2. What processes do experts use as they judge creativity in contest settings; are the processes involved in creativity contests similar to or different from those used in the experts' professional workplace settings?

3. What types of cognitive processing are involved in expert judgment of creative products and ideas; do experts use intuition, rational analysis, a combination of both, or something else when judging creativity contests?

4. What do experts say are the differences between how experts and non-experts make judgments of creativity; what kinds of mistakes might a novice make in judging creativity?

5. What biases and heuristics do expert judges acknowledge encountering when judging creativity in real-world situations, and how do they attempt to deal with them?

The following section discusses the findings and data collected in this research study and analyzes the findings as they relate to each specific research question. Contributions to theory and practice resulting from the research study are also discussed.

V.2 Discussion

V.2.1 *Definition and Elements of Creativity*

> I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description [of hard-core pornography]; and perhaps I could never succeed in intelligibly doing so. ***But I know it when I see it***, and the motion picture involved in this case is not that. Justice Potter Stewart, *Jacobellis v. Ohio*, 378 U.S. 184 (1964), *conc. op*. [emphasis added].

The question of how to define, identify and measure creativity has plagued researchers for more than 50 years, almost as long ago as when Justice Potter wrote the famous passage above categorizing obscenity as: "I know it when I see it." The results of this research suggest that, at least for domain experts, the answer to whether something is creative is similarly explained: they

simply know it when they see it although, by strict definition, they could not have seen it before because it did not exist, i.e., it had not been created. How domain experts know creativity when they "see it"—the criteria they apply and the processes they undertake to assess creativity—has no such ready answer.

The primary question for this research of 'how to define creativity' arose in response to concerns raised in literature suggesting that inconsistent and potentially contradictory results in creativity research might be due to a misalignment between the two-item construct definition used to describe creativity, on the one hand, and the unitary operational definition of creativity often used to measure the dependent variable of creative output. The results of this research study indicate that domain experts do not agree with either the two-item "novel and useful" construct or the unitary operational definition of creativity common in creativity research. In fact, participants specifically rejected the "novel and useful" definition of creativity widely adopted in research as being too simplistic, incomplete and inadequate in practice, and strongly questions the viability of that definition for research purposes.

In practice, the participants in this study reported using a minimum of three key factors to define, identify and measure professional creative output: 1) novelty; 2) inventiveness; and 3) task appropriateness. The weight of the data collected indicates novelty, inventiveness and task appropriateness are formative, elemental factors that are minimally necessary to identify and sufficiently measure the construct of creative output. Moreover, the participants identified three additional aspects they look for in the assessment and measurement of creativity, expecting to identify at least one of the following three criteria: 4) cleverness; 5) artisanship; and 6) execution. Some participants viewed the additional criteria as comprising the balance of six

formative elements of the creativity construct, while others viewed the additional aspects as second-order or perhaps reflective measures of the first three more formative factors.

Creativity domain experts also do not appear to use either intuition or rational analysis alone in reaching their judgments of creative output. Instead, participants reported and were observed applying a two-step or an iterative process: 1) a relatively immediate holistic recognition and binary decision that an item is or is not highly creative; and 2) either a very brief confirmatory analysis, if an item was determined to be not highly creative, or a more extensive analysis of highly creative items. In both cases, judges appeared to apply non-consciously the three-item criterion to all items assessed, and applied the multi-faceted six-factor measurement construct to assess the comparative level of creativity to more creative items, and award a final score.

In the first immediate assessment step, participants made it clear that novelty alone does not make an item creative: the item must represent a significant creative departure from existing products or ideas. This makes logical sense given that just because something is "new" or previously unseen does make it creative. It may be the rater has not seen the exact item before but recognizes it as an easily anticipated or a logical extension of an existing product or idea, i.e., "obvious". For example, adding a third hole to two-hole punched paper for the first time might be new and even novel compared to the state of the art of paper products at the time, but that alone does not make it creative. For an item to be considered creative, participants stated they required it to be novel *and* not an obvious extension of an existing product or idea; it must be unique, different, novel, and unexpected. In other words, it must represent an inventive step—enough of a departure from the current state of the art to surprise, astonish or challenge the expert judge's expectations. Lastly, more than mere "utility", experts reported a creative item must respond both appropriately and substantially to the creative task presented.

Once judges identified an item as exhibiting some level of creativity using the first three criteria, participants reported the amount of time and effort involved in determining whether and to what extent the item is highly creative depends on the complexity of the item, the degree of departure from the "state of the art", and the perceived difficulty in execution. This requires assessment of the cleverness of the item's idea or approach, the degree of artisanship exhibited by the finished product, and the elegance and impact of the execution. Thus, whether considered as additional factors of the main construct or as second-order reflective measures, most participants indicated that some or all three additional factors needed to be present to identify and measure an item as highly creative.

> "Creativity" is more than just being different. Anybody can play weird; that's easy. What's hard is to be as simple as Bach. Making the simple, awesomely simple, that's creativity.
>
> - Charles Mingus, Jazz composer, performer and pioneer (1922-1979).

For those participants who viewed the three additional measures as second-order or reflective factors, the elements appeared to help identify or qualify the main factors. For example, many participants indicated that cleverness could be viewed as either an extension or a more granular measure of novelty (where cleverness denotes a high degree of novelty) or inventiveness (as some degree of cleverness would be expected in an inventive step and, the more clever, the more inventive). Conversely, many of the participants noted that cleverness can be and often is a separately assessed element. For example, where two similar items are each novel and also to the same degree an inventive step forward in the state of the art, the degree of cleverness of each would allow an expert judge to distinguish between the two and rate one higher than the other. Likewise, artisanship and execution might be considered extensions of the concept of appropriateness to the task in some instances, although during the simulations and

task analysis exercises all respondents reported something much more significant than mere task utility must be present to identify an item as creative—the item must be both appropriate to the task and also elegantly and meaningfully deliver on the concept. As one participant put it: "If it's done technically correct, maybe it's perfect, but if falls flat on the messaging or it's boring, it doesn't communicate, then it shouldn't win an award." (David). Thus, it remains an open question whether artisanship and execution are additional formative factors necessary for a finding of creativity, or reflective measures used to measure the extent of achievement of task appropriateness.[3]

Interestingly, the six factors identified by the participants in this study closely resemble the elements of creativity suggested more than 50 years ago by Jackson & Messick (1965) in their four-item definition of creativity: 1) unusual when compared to other products; 2) appropriate to the context; 3) representing a shift in the constraints and boundaries of the situation; and 4) able to condense both simplicity and complexity. A minority of researchers have similarly argued for a three-factor definition of creativity that incorporates the elements of surprise, non-obviousness or inventiveness similar to the three-part test used by the Patent Offices of the United States and the United Kingdom to assess patent applications (Barton, 2003; Boden 2004; Simonton 2013).

---

[3] The following are other words participants used when attempting to describe each of the six key indicators, as generated by a word usage analysis drawn from interview transcripts and researcher observation notes:

Novel: unique, different, diverse, unusual, distinctive.

Inventive: surprising, unexpected, progressive, astonishing, unlikely, unforeseen, not obvious.

Appropriate: useful, functional, utility, fit-for-purpose, meaningful.

Clever: interesting, inspiring, appealing, compelling, conceptual, ideation, provocative, radical, exciting, extraordinary, daring, bold.

Artisanship: quality, commonsensical, beautiful, elegant, simplified, excellence, artistic, superior.

Execution: Delivery, achievement, challenge, success, realization, impactful.

The explanations provided by the participants in this study help to expand on those conceptual discussions by identifying and refining the factors suggested by researchers into the six factors identified as used in practice. The results of this research strongly suggest the definition of creativity, and the criteria used to assess creative outputs, needs to be expanded in research to include at a minimum a third criterion of non-obviousness, surprise or an inventive step. In addition, participants stated usefulness fails to capture the of utility sufficiently; appropriateness to the task allows for a more complete assessment, particularly in professional settings. Moreover, this research indicates that the other factors—cleverness, artisanship and execution—are either formative elements or reflective measures of creativity in practice, and that additional research is needed to determine if these factors are also necessary to more completely define and measure creativity in research.

V.2.2 *Processes Used by Experts in the Field*

The participants in this study reported they use almost exactly the same processes, criteria and standards to judge creativity in contest environments that they use in the workplace. One of the key differences between the two environments related to the time available to process an item and the evaluator's ability to put context around the item. Participants indicated the complexity of the items reviewed is one way the judgment process is significantly slowed down in a contest setting; more complex items require more time to evaluate and score. However, most participants reported that taking additional time to review an item rarely, if ever, caused the initial assessment or score to improve. In fact, most participants indicated additional time spent examining an item in a contest setting usually resulted in a lowering of the initial score. One possible explanation is that more time reviewing an item allowed the judge to be more critical, using a more analytic reasoning process to overcome initial intuitive impressions. There is no

evidence in this study, however, to indicate that analytic reasoning resulted in a more "correct" score, and it is likely judges avoided unnecessary analysis as a way to prevent bias from affecting their judgment. Moreover, not having as much information about a contest item as one might have about an item in a workplace setting was not seen as detrimental to the validity or reliability of assessments in contests. In fact, participants believed creativity is more easily identified when directly and contemporaneously assessed against multiple items—a situation that rarely occurs in practice.

There is also no evidence from this study to suggest the judges used any formulaic or arithmetic approach to the evaluation of creative output, other than the ranking effect of choosing a score to give an item. And there is no evidence to suggest expert judges require additional instruction or direction to use a specific criterion. In many of the studies reviewed prior to this research, researchers instructed participants to use either a single item construct (i.e., creative) or a two or three factor construct including novelty and usefulness, to assess creative products or ideas. However, this research clearly shows that experts do not expressly rely on novelty and usefulness alone, and do not appear to (and reported they did not) separately weight novelty, usefulness or any other criteria when evaluating creativity. This suggests experts assess creativity using a multi-dimensional holistic model to reach a scaled ultimate unitary decision without conscious processing or scoring of individual factors.

V.2.3 *Applicability of the Consensual Assessment Technique*

This research also supports Amabile's Consensual Assessment Technique (CAT) as a valid method in the process of evaluating creative output in both practice and research. Experts reported their assessment process followed the same general requirements outlined by Amabile for CAT evaluation (a group of domain experts, judging each item independently and without

specific instructions) and they were observed following those same processes in scoring items in the actual contest. The only instructions given the judges during the observed contest were an outline of the scoring scale to be used and specifics on how to enter scores electronically. CAT presumes domain experts have the requisite experience and understanding to recognize and assess creative output without pre-defined criteria or instructions, and if acting independently they will achieve a high level of agreement on the measure of creativity. Although CAT has been defined as using a "unitary" operational construct to measure creativity, in practice it appears domain experts applying CAT non-consciously rely instead on a multi-factor construct. Whether an item is creative, or more creative than another, is likely the final measure of the output's combined novelty, inventiveness, and appropriateness to the task, along with some measure of cleverness, artisanship and execution.

In summary, the evidence developed by this research suggests experts reach an initial binary conclusion (creative or not) and then assign a score reflecting a relative comparison of the level of creativity against other items either in the assessed set or against the expert's own experience and knowledge of the state of the art. In practice, this research also highlights the importance of using experienced professionals to identify creative output to determine which ideas and products to pursue, something business organizations find difficult to accomplish consistently (Harvey & Kou, 2013).

V.2.4 *Cognitive Processing (CEST and CCT)*

Participants in this study acknowledged that creativity triggers an immediate emotional response and, in many cases almost simultaneously, an abbreviated form of analysis occurs as part of the evaluator's assessment. Thus, the evidence generated by this research tends to support both Cognitive-Experiential Self-Theory (CEST) and Cognitive Continuum Theory

(CCT) in the context of how expertise-based intuition and rational analysis operates in evaluating creative output in practice. Both CEST and CCT recognize expertise-based intuition and rational analysis as components of evaluative cognition. However, CEST posits rational analysis and intuition operate either in direct opposition or at different times during the evaluation process. CCT, on the other hand, theorizes that intuition and rational analysis exist on a continuum and operate simultaneously, in varying respective degrees, in response to task stimuli. Both theories are supported by the findings of this research but the study failed to provide sufficient definitive evidence to differentiate between them, as each model adequately explains the results obtained.

As noted in the Findings chapter, participant judges appeared to use intuition for most of the initial assessments generated during the contest judging event. Moreover, all participants either acknowledged "intuition" was a critical component of their judgment process or reported experiencing several of the indicators of intuitive decision making such as using their "gut" instinct or reaching a decision with little time to fully consider each item. Participants also were observed moving very quickly through each entry, in most cases spending less than two seconds viewing each before entering a score. Participants also identified several aspects of their process consistent with theories of intuition, including sudden awareness of an idea or choice without conscious awareness of the source, affect-laden decisions often coupled with strong emotional or physical reactions, and high confidence in their judgments (many saying they rarely if ever go back to reconsider their decision). Other participants used the words "visceral reaction" to describe their judgments during the evaluation phase.

However, many of the participants also said they were on occasion concerned their fast and easily achieved decisions may not be accurate and that additional time and "cognitive effort" were sometimes required to determine whether their first "impressions" were accurate.

Participants reported that occasionally the additional time spent on review would uncover issues causing them to award a lower final score than first intended. None of the judges indicated that additional time spent on an entry resulted in a higher score. These findings suggest a two-step serial process as predicted by CEST whereby judges use the default mode of processing for the majority of decisions that match a particular item with a pattern or experience from long-term memory. However, when an entry is complex or initially considered particularly creative, a switch in processing can occur, causing the participant to apply the slower, more cognitively involved effort of rational thought.

The findings also support the predictions of CCT that in many cases intuition and rational thinking occur simultaneously as "quasirationality" in response to the demands of complex decision tasks. Entering scores for each item often took judges longer than the time required to review the item, and many items appeared to be scored "in bulk" for this reason. However, in some cases, particularly when an item contained multiple parts or dimensions, judges spent additional time looking at the entry more holistically, in an apparent attempt to take in the overall concept or execution of the item. In most of these cases, judges still moved very quickly and deliberately, reviewing each item, scoring them and moving on. Judges reported that items receiving low scores (5 or less) were reviewed for the least amount of time, while items that received high scores (8 or above) were not only reviewed for a longer period of time (ten or more seconds), but often judges were seen contemplating the item from multiple perspectives. Judges later explained the additional time was often spent "admiring" a particularly creative item they had already decided would receive a very high (9 or10) score. When asked to explain their overall scoring system and the criteria they were using in these specific instances, participants could not verbalize their reasoning saying they would have to "dumb it down" to an elementary

explanation. However, all participants agreed the process used in these cases did not involve conscious switching or toggling from one processing mode to another. In fact, many judges reported their process involved both a holistic appreciation for the item, in tandem with a more critical appreciation for discrete elements of the execution and approach, and that their thought process involved "searching" for the right response to the item presented. This suggests judges were not operating solely in one mode or the other, as predicted by CEST, but were moving along a continuum between rational analysis and emotional intuitive response to reach a decision that matched the stimulus, as theorized by CCT.

Evidence from this research also suggests that the two theories may describe different aspects of the same overall process. During observations, judges exhibited signs suggesting they may have been operating in a binary "off-on" approach, using intuition to quickly assess whether an item deserved further consideration. When an item elicited an intuitive response, judges appeared to then shift into a more holistic review of the item, but not into a purely analytical review. Participants reported that even when they identified an item as creative, the initial emotional response did not disappear as they continued to review the item in more detail, and that they did not recall critically analyzing an item unless intuitively they felt something wasn't quite complete. In many instances reported by the participants and observed during the judging and simulations, judges were unable to identify why a particular item was considered highly creative, but very often they could readily identify why an item might *not* receive a higher score. This suggests judges were using a continuum approach to processing along with a form of pattern matching. Judges first compared the item against their memory and observed immediately if the item matched their past experience (i.e., was *not* new or different) requiring no movement along the continuum (due to having reached a successful decision outcome) and

they awarded a low or middle score. If the item did not match the judge's past experience or memory, a shift to a different processing approach was required to consider the item further (due to decision failure). All of the participants indicated a more complex form of decision-making was involved in this "second look." However, none of the judges exhibited and none of the participants reported that they toggled into a purely analytic mode.

This task-centered response and movement along a continuum until an acceptable choice is found supports CCT's quasirationailty movement model, but does not exclude the CCT model. In fact, most participants indicated they had used both approaches, at different times or for different categories, suggesting both theories may adequately describe the process of creative evaluation and that the research methodology employed here was insufficient to identify the boundary conditions of either. Likewise, it could be that both processes are available to experts and are used for different purposes or in different contexts to assess creativity. "If there are two models that both agree with observation…then one cannot say that one is more real than another." (Hawking & Mlodinow, 2010: 46).

V.2.5 *Recognition Primed Decision-Making and Expert Intuition*

Regardless of whether CEST or CCT, or both, are involved, the evidence strongly supports experts utilize one or more of the variants of the Recognition Primed Decision (RPD) model when assessing creative products and ideas. Interestingly, the participants appeared to use the opposite of "pattern-matching" and "consistency with prior experience" that are the hallmarks of most variants of RPD. For creative products, most participants reported they evaluated entries very quickly, using both intuition and an abbreviated analysis to determine that the expert had *not* seen the item, i.e., that the item was *outside* of the expert's stored memory or recent

experience, before assessing the qualities of novelty and cleverness. In short, the participants appeared to look for what did not fit a pattern or was inconsistent with expectation.

Participants then appeared to use story-building to simulate and evaluate the creative steps that might have been taken to develop the item (to assess inventiveness, artisanship and execution), and lastly participants incorporated *forward-chain reasoning* to extend the simulation into future scenarios to evaluate the likelihood of success of the product or idea (to assess functionality, task appropriateness, and concept delivery). Highly creative items would then require "progressive deepening" of simulation and comparison, and extended imagination, in order to complete forward chain reasoning. By definition, "forward-chaining" requires assuming some aspects of an item as a given (the "if") and extending those aspects into the future until an imagined solution is achieved (the "then"). Forward chaining is not rational analysis where various "options" are considered against each other using logic to derive a superior solution. In RPD and forward-chaining used by experts, only one "option" is simulated, using imagination in an attempt to recreate the solution presented.

In this research study, all participants appeared to utilize some variant of RPD to identify and assess items they had intuitively determined were creative and worthy of further assessment. This evidence supports Klein's model of RPD as being "a blend of intuition and analysis. The pattern matching is the intuitive part, and the mental simulation is the conscious, deliberate, and analytical part." (Klein, 2008: 458). In a fashion, RPD helps to explain how both CEST and CCT may be involved in cognitive processing in uncertain, ambiguous, time-sensitive and non-algorithmic tasks. As pointed out by other researchers, the process of "intuiting" is rapid (often instantaneous), spontaneous (without effort and unable to be controlled) and alogical (not necessarily contradicting the rules of logic but may not follow them either) (Dorfler &

Ackermann, 2012). Intuition is ineffective in decision-making involving algorithmic tasks that can be decomposed and solved logically, sequentially, or mathematically but can be highly effective in non-algorithmic tasks (Dane, Rockmann & Pratt, 2012). RPD is also less effective in situations presenting highly combinatorial or algorithmic problems, where justifications are required, or in cases where differing views of multiple stakeholders must be considered (Klein, 1998). In this study, the quantitative analysis of judges' scores also supports the conclusion that domain experts with significant experience (from "high-fidelity" environments) utilizing expert intuition and consensual assessment techniques achieve more reliable and valid results in non-algorithmic tasks than non-experts. These results also might explain why organizations that assess creativity in group settings make less effective decisions—whenever justifications are required or multiple stakeholder views must be considered, expert intuition is likely to be curtailed or ineffective.

V.2.6 *Expert v. Novice*

All of the participants in this study expressed the opinion that novices would struggle to consistently identify and score highly creative work. Participants identified several reasons for the challenges novices would face, including: a) an incomplete awareness of the "state of the art" (lack of domain familiarity); b) the absence of understanding the work effort involved in the varying levels of artisanship required to produce highly creative work (lack of creative experience); c) the relatively small amount of judging experience (lack of discernment skill), and d) the lack of experience with heuristics and biases (lack of debiasing skill) that would result in personal subjectivity and emotional responses going unanalyzed. In essence, participants believed novices would unrealistically assume their abilities to assess creativity were greater than their experience would allow and succumb to untrained emotional responses. Some participants

termed this gap of knowledge and experience the "novice trap" or referred to the novice's "inability to know what they don't know." In psychology, this cognitive bias is termed the Dunning-Kruger Effect and explains the difficulty inexperienced individuals have in recognizing their own lack of knowledge and the overconfidence in decision making that often results (Dunning, et al., 2003).

Participants also described several ways the challenges would likely inhibit the ability of a novice judge to assess creativity effectively. First and foremost, the participants identified the risk of "premature convergence" or accepting the most easily identified response and failing to see additional potential solutions. Premature convergence is more likely in tasks with short-time frames and in complex environments, particularly where the decision-maker lacks experience in the task and has limited breadth of understanding of the field. For novice judges, the time pressures and lack of experience creates the risk they would seek out only items that appear very different from the rest of the field and select those items as highly creative without further evaluation. Participants also indicated that a lack of experience would cause novices to not recognize "rough creativity", i.e., creative work that is not fully polished but which exhibits all of the traits of creativity except execution and artisanship. One example of this challenge in practice would be highly creative formative work that needs further development and which is so divergent a novice would assume it would be impossible to achieve. Experts, on the other hand, would be able to identify the potentiality of such rough creativity and understand how it might be implemented in the future.

The participants not only identified the necessity of using domain experts to ensure the validity and reliability of creativity assessment, the experts' descriptions of the processes and criteria used to identify highly creative work in contest settings also reveals some of the reasons

behind this necessity. Faced with the significant time pressures, uncertainty and ambiguity of a high-stakes creativity contest, experts must rely on various heuristics in order to accomplish the tasks timely. However, the use of heuristics also creates the potential for the impact of numerous biases to alter the results. The methods used by experts as expressed by the participants in this study not only serve to accomplish a great deal in a short period of time, but as explained in the next section, also help to reduce the frequency and the impact of bias on the decisions made. Non-experts lack the experience necessary to appropriately utilize heuristics and to identify and avoid biases that would result.

V.2.7 *Using Heuristics and Avoiding Biases*

Participants in this study acknowledged using heuristics to judge creativity in high-pressure, low-time environments, and all acknowledged numerous potential biases that could adversely affect their judgment and decision-making. Participants identified a number of biases novices would encounter in judging creativity that they would struggle to recognize and deal with effectively, including: availability, typicality, anchoring, and confirmation bias. Novices without experience in the relevant domain, lacking awareness of the processes and criteria necessary for valid assessment, and inexperienced in the process of judging and preventing biases from affecting their decisions, are ill equipped to provide reliable judgments in the face of these challenges. However, the participants reported several techniques they relied upon to reduce the likelihood and impact of biases.

Several participants identified “mindfulness” as a significant method used to avoid the impact of biases or predisposition on decisions, by having full consciousness of the situation, the importance of the work to the contestants, and the value of the contribution to the overall profession of creativity. As one participant explained:

> "These contests are really important to the participants, and I have to do my very best to make good decisions and not let my own subjective opinions or the feelings of others affect my judgment. I approach these [contests] with a high degree of awareness of the ultimate goal is: identifying the most creative and worthwhile work." [Edward]

Other participants indicated they felt that prior contest judging and professional work experience had improved their ability to avoid certain biases from creeping into their decisions. For many, the same experience gained in the "high-fidelity" environment of a contest, coupled with the immediate feedback of judging professional work, in which they developed expertise in the domain also trained them to recognize bias and debias their judgments. Indeed, training and expertise gained from practical experience, along with having empathy for the individuals the decision will impact, have been shown to improve "debiasing" in decision-making (Dhami, 2013).

Participants also pointed out the importance of staying abreast of other creativity awards contest outcomes and the creative work published in their respective industries as well as other fields as a means of avoiding subjectivity and other potential biases. Many participants specifically reported that critically assessing the creative work of others outside of the participant's own professional work and contests in which they acted as judges was a way of keeping an open mind and avoiding "group-think."

While participants acknowledged that time pressures often required the use of heuristics to complete assessment of creativity in practice that could lead to bias in judgment, most felt "mindfulness" of the potential biases, awareness of their personal subjectivities, and use of some form of analysis to confirm or discount initial reactions, all helped experienced judges avoid or reduce the likelihood of incorrect biases in their judgments. Kahneman acknowledged that

intuition can produce valid and accurate judgments and decisions but only in situations where domain expertise overcomes inherent biases caused by heuristics (Kahneman, 2011). This study supports the concept that domain expertise can overcome inherent biases where that expertise is obtained and maintained in "high-fidelity environments," and where the judgments and decisions surround tasks that are non-algorithmic.

V.2.8 *Inter-rater Agreement of Judges Scores*

As noted in the Findings section, inter-rater correlation of judging scores across eight contests over multiple years were calculated at a low of 0.674 and a high of .797. ICC scores of reliability between .65 and .80 are considered to reflect a moderate to high degree of agreement in creativity research. However, considering the level of domain expertise and relative homogeny of items assessed in the creativity contest, the judges' interrater correlations could be criticized for not being even higher. As pointed out in the literature review, fewer numbers of experts have been shown to achieve a high level of inter-rater agreement compared to the number of novices necessary to achieve a similar rating (Kaufman, et al., 2008). However, assessing creativity by its nature is a human-based non-algorithmic process with no clear objective determinant, and therefore anything nearing total agreement among multiple judges is unachievable. In fact, had the correlation of the judges' scores been higher than .80, the research results would be suspect for being too highly correlated. All of the work assessed was at a very high professional level—presumably the contest entries had been pre-selected as highly creative by the entrants, all of whom were creative professionals. Despite this, judges' scores on individual entries ranged from a low of 0 to a high of 10. Given that experienced professionals created the work judged in a field setting, it would be unlikely, and highly suspect, if the judges' scores achieved a very high level of agreement. To achieve very high correlations one would

expect only one or a few items in each category to be scored as highly creative with the remaining items scoring well below the top items along a normal distribution, allowing for complete discrimination across the top, middle and bottom ranges. In situations where professional work is being judged, however, less variation would be expected among the items scored with clustering near the very top of the range, making high inter-rater agreement much more difficult to achieve. Accordingly, score correlations above .70 for highly creative products across a small number of raters would not be expected unless the assessments were in fact highly reliable.

V.3 Contributions to Theory And Practice

This field study contributes to the development of creativity research and the understanding of process models of judgment and decision-making in several ways. First, through the observation and in-depth inquiry of the processes and criteria that expert judge participants used in assessing creative entries, this research develops evidence highlighting the critical role experts must play in the measurement of creative output in research. The experts observed in this research revealed unique and special processes and approaches to the evaluation of creative products that appear to be very fast and highly effective, without the need for special instruction or significant support. Moreover, the participants' judgments of the entries appeared to be highly valid as shown by the tight consensus on the choices of winning entries and highly reliable as shown by the results of inter-rater reliability testing on the participants' actual scoring and the scores given by judges in prior contests.

Second, participants identified several deficiencies in the operational and construct definitions of creativity often used in creativity research. Participants provided evidence suggesting that a holistic unitary measure of creativity, applying multiple facets that may be

different depending on the item assessed, is more effective for experts in assessing creative products in practice. Because of their years of experience evaluating creative products and ideas, experts appear to be able to identify the key components of creativity for a particular item or, perhaps more correctly, identify when key elements are missing or poorly executed. While it is tempting to seize upon the key elements and components identified by the participants as indicative of creativity in order to create a list for potential use by non-experts, it is important to remember that this study involved only a small number of participant judges in the think-aloud and simulation exercises that generated those key elements identified. And each of the participants made clear that judging a different group of items would likely involve different combinations of elements and components, such that not all should be considered necessary or important for every creative product or idea. Moreover, none of the experts felt non-experts would be able to apply specific criteria. It was only by means of the expert participants' experience and deliberate practice that they had learned what to look for in which situations, and what allows experts to "know it when they see it."

Third, this research potentially highlights some of the context and reasons behind quantitative research showing the stark differences in performance between experts and non-experts when judging creativity. Studies have shown that non-experts' assessments of creativity in small groups cannot be considered reliable or valid, and that it can take as many as a 100 non-experts to achieve an inter-rater reliability equal to just a few experts. This research provides additional evidence of the challenges inherent in using non-experts by identifying some of the reasons non-experts would struggle to evaluate creativity and achieve reliable results.

The participants' behavior and processes also contribute to understanding the cognitive processing of experts during decision making in general, and of Cognitive-Experiential Self-

Theory and Cognitive Continuum Theory in particular. CEST and CCT both recognize the existence and effective use of intuition by experts, but the mechanisms by which this occurs are not well understood. The results of this study provide insights into expert decision-making behavior and their perceptions about evaluating creative products in a field setting.

This field study also contributes to practice by identifying ways to improve the assessment of creativity in organizations. Managers of organizations struggle to differentiate creativity, innovation or problem solving, often incorrectly viewing the three separate constructs as inextricably connected (Banks, et al., 2002). Organizational groups also tend to identify relatively average ideas produced by their members as more creative than other more novel ideas created by other individuals (Rietzschel, et al., 2006). The findings of this study highlight the importance of employing domain experts to evaluate creative products and ideas and reveal that assessment of creativity is not a team sport. As required by the Consensual Assessment Technique, and recommended by the participants in this study, assessment of creativity should be done by individuals acting independently and without specific instructions. Using groups to assess creativity may be problematic to the extent it introduces potential bias from other participants, including confirmatory bias and group-think, as well as political effects as members "choose consciously or subconsciously to ignore ideas, advocate for their own ideas, show enthusiasm for others' ideas, and provide interpersonal rewards for good ideas" (Rietzschel et al., 2006: 347).

VI CONCLUSIONS AND RECOMMENDATIONS

VI.1 Conclusions

This research provides rich insight into how small groups of domain experts identify and evaluate the creative output of professionals in the real world setting of an awards contest. The results of the research contradict some of the limited models and measurement constructs of creativity used in research, and challenge the validity and reliability of research studies that relied on non-experts to assess creative output. The results support the value of expert intuition in making rapid, valid and reliable assessments of creativity, and the limited role rational analysis plays in quantifying creativity of professional output and in reducing judgment error from decision biases. While the results of the research cannot be generalized to a wider population, the observations, analysis and reports from the participants provide strong evidence in support of a broader operational definition and multiple factor measures of creativity in products and ideas for both research and practice. Lastly, this research suggests numerous opportunities for further research to refine our understanding of the nature of creativity and identify how to better measure creative output.

VI.2 Recommendations for further research

Quantitative data in this study indicates that many of the expert judges of the creativity contests reviewed did not award scores using the entire scale proposed by the contest administrators. A review of the minimum and maximum scores given by judges in the contests of this study indicated that, while most judges used the proposed 1-10 scale, some judges used a scale ranging from 2 to 10, others used a scale of 5 to 9, and several judges used a scale of 0 to 9 but gave no entries a top scores of 10. Future research using detailed statistical analyses might provide insights into whether rater effects affected the results of the scores or the contest. Rater

effect analysis might also reveal whether expert judge scoring is less reliable in the "mid-range" of creativity where the bias of subjectivity and personal variations in the weighting of various factors of the creativity concept might have a greater impact. Another line of research might involve test-retest reliability of expert judges over a long period. Would an item be considered as creative in the years following its initial assessment? If so, what role if any does the criterion of novelty play in the decision? As one of the criteria used by most judges is whether an entry is "different," "new", or "surprising", it is difficult to anticipate the consistency of judges' scores over time when that characteristic is no longer apparent. However, an experiment using similar or the same items from a prior contest that subsequently asks judges to score them as to the level of creativity at the time the item was first scored may yield insights into whether judges are consistent in the comparison of items after the passage of time.

Brain imaging studies using fMRI have shown a correlation between the areas of the brain engaged for different types of social and mechanical tasks (see Jack, et al, 2014) and suggest that when the positive task system is engaged the default system is deactivated. A similar study using both simple and complex items previously judged by experts as highly creative and less creative might reveal insights into whether different information processing systems operate in parallel or serial fashion when making aesthetic judgments.

Future research might also investigate whether expert and non-expert judges produce more or less valid and reliable results when evaluating creativity in a group as opposed to independently. The results of this research confirmed that awards judging in practice follows the requirements set out in Amabile's Consensual Assessment Technique that evaluators must make their judgments independently and without consultation. One assumption of CAT appears to be that judges acting together would negatively influence each other's judgment. However, several

participants in this study reported that judgments for “Best in Class” winners were made in open group discussions and the judges expressed that the ultimate choices made in that setting were highly valid. Additional research could test these assumptions among both non-experts and experts, to see if non-expert judgments can be improved when evaluating creativity collaboratively, and to assess whether expert judgment is less reliable when made in a group setting as suggested by research into organizational creativity assessments. Research on the latter question would have significant implications on the assessment of creativity in practice, potentially leading to better models of decision making in organizations.

Similarly, a factor analysis comparing judges scores by separately applying each of the six key elements of creativity reported in this study might reveal significant variation in what judges rely on for their conclusions while still awarding very similar overall scores. For example, one judge might give an entry high marks for novelty and execution, and thus a high overall score where those items are heavily weighted, in comparison to another judge who might heavily weight inventiveness and artisanship and score an item as high in creativity for quite different reasons.

VI.3 Final Thoughts

Obviously, it would be simpler and cheaper if researchers could employ non-experts or untrained evaluators to score creativity using an objective multi-dimension scale and achieve relatively high validity and reliability. However, this research and other studies indicate non-experts are ineffective judges of creativity and that objective scales are not used by domain experts. Conversely, requiring the use of domain expert judges for every research study might be overkill, particularly for simple evaluations of basic creativity. For example, Amabile, who developed the Consensual Assessment Technique, successfully used psychology students to

evaluate collages created by elementary school students in her research of task motivation. Such an approach might be perfectly appropriate in some contexts; nonetheless, this study shows that researchers need to use greater care when designing and evaluating creativity research studies. Employing domain experts with experience in judging creativity and applying the appropriate criteria is critical to assure reliability and validity in studies of creative outcome.

APPENDICES

Appendix A: Applied Cognitive Task Analysis

This research utilized a modified form of cognitive task analysis called "applied cognitive task analysis" to conduct semi-structured probe question interviews of expert judges. Cognitive task analysis (CTA) is a set of methods designed to describe the various cognitive skills and mental demands required to accomplish challenging and complex decisions and judgments (Crandell, Klein & Hoffman, 2006). While there are many different types of cognitive analysis methods and approaches within CTA, each aimed at achieving a unique aspect of cognitive research, most are resource intensive, difficult to use, and often require specialized training or experience in cognitive psychology (Hoffman & Woods, 2000). As a result, many CTA methods tend to be of limited use to management scholars and researchers (Militello & Hutton, 1998).

Applied cognitive task analysis (ACTA) is a streamlined version of cognitive task analysis developed for use by researchers not trained in cognitive psychology (Militello & Hutton 1998; Crandell, Klein & Hoffman, 2006). ACTA provides a refined set of cognitive task analysis techniques specifically designed to identify the key cognitive elements required to perform mentally challenging or complex tasks (McAndrew & Gore 2013). Because ACTA focuses on the key cognitive elements underlying difficult judgments and decisions, critical cues and patterns and problem-solving strategies, it is particularly well suited to studying expert decision making. ACTA has been used successfully in studies involving weather forecasting (Hoffman et al., 2006), clinical nursing (Militellto & Lim 1995), currency trading (McAndrew & Gore, 2013), military command and control (Drury & Darling, 2008), recruitment (Gore & Riley, 2004) and financial markets (McAndrew & Gore, 2007).

As a streamlined method, ACTA presents a likely trade-off between usability and resources, on the one hand, and power and comprehensiveness, on the other (Militello & Hutton, 1998; Crandell, Klein & Hoffman, 2006). As a result, ACTA techniques should not be expected to produce information as comprehensive and specific as other more in-depth and systematic forms of cognitive task analysis. However, ACTA methods have been shown to consistently produce reliable, high-quality information about cognitive processes in various contexts, as well as being more accessible to researchers without intensive training in cognitive psychology (Militello & Hutton, 1998; McAndrew & Gore, 2013).

The ACTA Process:

The process of applied cognitive task analysis involves four complementary steps that are designed to systematically build on one another to elicit high quality, task specific knowledge (McAndrew & Gore, 2013). ACTA utilizes structured interviews, observation, and simulation to elicit knowledge from subject matter experts (SMEs), and knowledge representation techniques, e.g., cognitive mapping, to provide structure for organizing and comparing cognitive information. The first step of ACTA involves the interviewer and the SME co-producing a task diagram providing a broad overview of the task. In this initial step of the interview, the SME is asked to decompose the task into subtasks with questions such as, “Think about what you do when you (task of interest). Can you break this task down into less than six but not less than three steps?” (Militello & Hutton, 1998). The SME is then asked which of the identified subtasks require difficult cognitive effort. This step provides a “surface level” view of the task and identifies specific areas of the task, or subtasks, requiring complex cognitive skills to be explored in greater detail in subsequent steps.

Step two of the process is the "knowledge audit," where the SME reviews and explicates the aspects of expertise required for effective execution of the critical cognitive subtasks identified in step one. The knowledge audit draws directly on research on expert-novice differences and critical decision method studies of expert decision-making to uncover the demands of cognitively challenging tasks (Militello & Hutton, 1998). During the audit, the interviewer uses generic and specific probes to elicit greater detail about the subtasks and cognitive effort necessary to complete them effectively, including critical cues and strategies for decision making. Concrete examples and specific information about past experiences and comparisons between how experts and novices might perform the task are captured with the goal of streamlining and improving data collection and analysis (Militello & Hutton, 1998; McAndrew & Gore, 2013). The output of the knowledge audit is an inventory of the task-specific expertise with examples of situations in which expertise is employed, cues and strategies that are used to make decisions and an explanation of why the decision poses a challenge to novices (Militello & Hutton, 1998).

The third step, the simulation interview, builds on the information obtained in the first two steps to contextualize the task, and allows the interviewer to better understand the SME's cognitive processes. In the simulation interview, the SME is presented with a challenging brief scenario that has been created in advance for the purposes of the interview. The interviewer presents the challenge and asks the SME to imagine going through the scenario, visualizing the steps involved. The interviewer then probes the SME on issues relating to situation assessment, potential errors and biases, cues and patterns and other challenges the situation might present (Militello & Hutton 1998). The information solicited is captured and recorded in a simulation interview table for subsequent comparison and analysis across interviews.

After the interviews are complete, the final step is creation of a "cognitive demands table" of the information elicited in all the interviews, to merge and synthesize the data. The cognitive demands table provides a format for the researcher to identify those areas requiring complex cognitive skills and the pertinent problem-solving and decision-making activities involved in the task (Militello & Hutton, 1998).

Appendix B: Review of Selected Creativity Studies

Author(s)	Year	Title	Publication	Subject types	Rating by	Inter-rater reliability
Akinola & Mendes	2008	The Dark Side of Creativity: Biological Vulnerability and Negative Emotions Lead to Greater Artistic Creativity	Personality & Social Psychology Bulletin	Young adults ages 18-25, general pop., N= 90; 65 females	4 "professional artists" and 4 studio art grad students	.65 to .88
Amabile	1982	Children's artistic creativity: Detrimental effects of competition in a field setting	Personality & Social Psychology Bulletin	Schoolgirls, N=22 ages 7-11	"artist-judges"	"high"
Amabile, et al	1986	Social influences on creativity: The effects of	Journal of Personality and Social Psychology	Study 1: N=115 boys and girls, ages 5-10; Study 2: N=80, students, ages 8-11; Study	Study 1: 3 elementary school teachers; Study 2: "artist-judges" nfs;	Study 1: .91; Study 2: .80; Study 3: .75

		contracted-for reward		3: N=60 undergrad women psych students	Study 3: 14 “artists” nsf	
Baer	1998	Gender differences in the effects of extrinsic motivation on creativity.	The Journal of creative behavior	N= 70 middle school students, 35 male and 35 female	4 “art educators”	0.8
Butler	1987	Task-involving and ego-involving properties of evaluation	Journal of education psychology	N=100, top performing 5th and 6th grade students	Count of # of original responses by 2 “judges”	.91-.93
Conti et al.	2001	The impact of competition on intrinsic motivation and creativity	Personality and Individual Differences	N=50, children ages 6-10	5 judges with “experience in children’s art”	0.82
Eisenberger & Armeli	1997	Can Salient Reward Increase Creative Performance Without Reducing Intrinsic Creative Interest?	Journal of Personality and Social Psychology	Study 1: N=296 5th and 6th grade students; Study 2: N=120 5th and 6th grade students	2 judges: two judges “assigned each drawing a score equal to the total number of times the same topic appeared in the population of drawings”	Study 1: 0.99; Study 2: .98

Eisenberger et al.	1998	Can the promise of reward increase creativity?	Journal of Personality and Social Psychology	Study 1: N=216 5th grade students; Study 2: 220 5th and 6th grade	2 judges, same as above; in Study 2, 2nd judge only rated 60 items	Study 1: .97; Study 2: .99 (60 only)
Eisenberger et al.	1999	Promised reward and creativity: Effects of prior experience	Journal of Experimental Social Psychology	N=238 5th and 6th graders	2 judges, same as above	0.99
Eisenberger & Rhoades	2001	Incremental effects of reward on creativity	Journal of Personality and Social Psychology	Study 1: N= 72 5th and 6th graders; Study 2: N= 97 5th graders ; Study 3: N+ ; Study 4: N=	Studies 1 & 2: 3 "judges," based on novelty combined with quality; Study 3: 3 undergrad research assistants; Study 4 N/A	Study 1 & 2: 0.88; Study 3: .82; Study 4 N/A
Eisenberger & Aselage	2009	Incremental effects of reward on experienced performance pressure: positive outcomes for intrinsic interest and creativity	Journal of Organizational Behavior	Study 3: N=405 intro psych students	2 undergraduate research assistants	"The intraclass correlation coefficient for the judges' creativity ratings was .64."
Friedman	2009	Reinvestigating the effects of promised reward on creativity	Creativity Research Journal	Study 1: N=81 undergrad intro psych students; Study 2: N=108	Study 1: 22 undergrad coders; Study 2: 14 coders	Study 1: .84; Study 2: .96

				undergrad intro psych students		
Gerrard, et al.	1996	Promoting children's creativity: Effects of competition, self-esteem, and immunization.	Creativity Research Journal	N=103 3rd grade children	2 art professors and 1 grad art student rated on “creativity” and “quality”; 21 teachers of gifted children also rated on “creativity”	.69-.72 for art judges; for teachers .89-.92
Glover & Zimmer	1982	Procedures to influence levels of questions asked by students.	The Journal of General Psychology	N=24 5th graders	2 ed psych students rated the quality of questions 5th grade students asked during class before and after treatment	Not given.
Hennessey	1989	The effect of extrinsic constraints on children's creativity while using a computer	Creativity Research Journal	N=66 children, age 7 - 13	Unknown	
Hennessey et al.	1989	Immunizing children against the negative effects of reward	Contemporary educational psychology	Study 1: N=113 3rd-5th graders, age 7-11; Study 2: N=58 3rd graders	Study 1; 3 elementary school teachers rated children's stories on “creativity”; Study 2: 12	Study 1: .80; Study 2: .70

					elementary school teachers "familiar with the work of 3rd graders" rated collages	
Hennessey & Zbikowski	1993	Immunizing children against the negative effects of reward: A further examination of intrinsic motivation training techniques	Creativity Research Journal	N=41 8-10 year olds	Rated creative stories. *Abstract only, no other info*	
Kachelmeier, et al.	2008	Measuring and Motivating Quantity, Creativity, or Both	Journal of Accounting Research	N=78 undergrad business students	11 doctoral students	CA .86
Moran & Liou	1982	Effects of reward on creativity in college students of two levels of ability.	Perceptual and motor skills	N=80 college students	2 "judges"	.85-.98
Selart et al.	2008	Effects of Reward on Self-regulation, Intrinsic Motivation and Creativity	Scandinavian Journal of Educational Research	N=42 psych undergrads	3 grad psych students	0.62

Aime, et al.	2014	The riddle of hierarchy: Power transitions in cross-functional teams	Academy of Management Journal	N=131, divided into teams of 4-5, business school students, avg age 21	3 upper level PhD. students with prior work experience. Scale 1 -10.	(ICC(1, k) = .63, ICC(2, k) = .65, F = 2.88, p< .001 , rwg = .82).
Chua	2013	The costs of ambient cultural disharmony: Indirect intercultural conflicts in social environment undermine creativity	Academy of Management Journal	Study 1: N=188, avg age 34, from Mturk; Study 2: N=264 college students	Study 1: 2 coders experienced in fashion design; ideas task; scale 1-7; Study 2: 2 entrepreneurs; business ideas task; scale 1-7	Study 1: ICC .78-.88, agreement .68-.81; Study 2: ICC .89-.90, agreement .82
Mattern, et al	2013	Matching Creativity Perceptions and Capabilities: Exploring the Impact of Feedback Messages.	Journal of Advertising Education	N=849 college students	4 researchers rated Alternative Uses Task responses for originality, flexibility and elaboration.	Cohen's Kappa: .92, .98 and .99
Schuhmacher and Kuester	2012	Identification of Lead User Characteristics Driving the Quality of Service Innovation Ideas	Creativity and Innovation Management	N=120 Mturk users, avg age 33 years	2 soccer club webpage user/managers, judged novelty, feasibility and relevance each on 7 point scale	Cohen's Kappa: .87

Grant & Berry	2011	The Necessity of Others is the Mother of Invention: Intrinsic and Prosocial Motivations, Perspective Taking, and Creativity	Academy of Management Journal	N=100 undergrads	Students with work experience in music business (3-4 years) rated ideas for generating revenue on creativity, 1-7 scale	ICC2: .69, agreement: .63
Santanen, et al.	2004	Causal Relationships in Creative Problem Solving: Comparing Facilitation In...	Journal of Management Information Systems	N=244 MIS undergrad students in 61 four person teams	6 disaster relief experts rated solutions to water crisis scenario; 4 university officials rated solutions for School of Business problem. 15 years avg experience	.834 for water crisis; .91 for business problem.
Ray & Romano	2013	Creative Problem Solving in GSS Groups: Do Creative Styles Matter?	Group Decision and Negotiation	N=250 business school students, avg age 28	Unstated number of judges (faculty with background in creativity and student judges) rated group ideas on Coffee Shop problem on novelty, cost-effectiveness	Unk

					and feasibility	
Shalley	1995	Effects of coaction, expected evaluation, and goal setting on creativity and productivity	Academy of Management Journal	N=84 undergrads, avg age 22	3 doctoral students with some HR work experience and MBA or MA judged creativity of responses to an HR manager in-basket complex-heuristic exercise, scale 1-7	Cronbach's alpha: .77
Perry-Smith & Shalley	2014	A Social Composition View of Team Creativity: The Role of Member Nationality-Heterogeneous Ties Outside of the Team	Organization science	N=82 long-term MBA teams of four to six individuals	2 doctoral students and 2 professors independently rated creativity of team's final projects	Interrater reliability: rwg2 with a mean rwg2 .82 and a median rwg2 .88.
Litchfield, et al.	2011	Directing idea generation using brainstorming with specific novelty goals	Motivation & Emotion	N=147 college freshman	2 college students rated novelty, creativity, effectiveness and practicality, scale 1-5	average rwg for novelty (.78), creativity (.79), effectiveness (.77), and practicality (.71) (James et

						al. 1984, 1993)
Binnewies, et al.	2008	Age and creativity at work: The interplay between job resources, age and idea creativity	Journal of Managerial Psychology	N=119 nurses who reported "creative ideas" in a broader survey, avg age 34	3 registered nurses with 10 to 30 years teaching experience, rated on novelty, usefulness and overall creativity, 1-10 scale	ICC of .89
Wynder	2007	The Interaction Between Domain-Relevant Knowledge and Control System Design on Creativity	Australian Journal of Management	N=63 college students avg age 20, provided ideas for a business problem	3 graduate business students with some work experience	Interrater reliability of .59-.65; Cronbach's alpha of .79.
Matthing, et al.	2006	Developing successful technology-based services: the issue of identifying and involving innovative users	The Journal of Services Marketing	N=52 university students and others on campus, provided new mobile phone service ideas	4 panels, 3 judges each from R&D, Tech and Marketing dept of phone company, plus a panel of 6 consumers, rated "originality" of ideas on 1-10 scale	Pearson's r, .69-.79, p<.01
Williams	2004	Personality, attitude, and leader influences	European Journal of	N=208 nonacademic employees of a university,	4 judges with graduate degrees in	ICC: novelty .71, usefulnes

		on divergent thinking and creativity in organizations	Innovation Management	asked to provide "suggestions for improvement"	management, experience working at universities and as management consultants, rated novelty and usefulness on 1-11 scale	s .59, combined as creativity .78
Marakas & Elam	1997	Creativity enhancement in problem solving: Through software or process?	Management Science	N=40 systems professionals from a local information center and senior undergrad and grad MIS students, asked to respond to open-ended questions	3 judges, members of faculty at business school	0.77
Sosak, et al.	1997	Effects of leadership style and anonymity on group potency and effectiveness in a group decision support system environment	Journal of Applied Psychology	N=159 undergrad students, asked to generate recommendations for an economic development program	2 "experts" rated on imaginativeness, innovativeness and value addition.	Cronbach's alpha: .90, .95, .97.
Runco, et al.	1994	Judgments of the creativity	The Journal of Psychology	N=47 visual art college students,	3 professional artist / art	not reported

		of artwork from students and professional artists		created a 3D art project	instructor judges, 1-7 scale, and students judged each other's work	
Mottweiler and Taylor	2014	Elaborated Role Play and Creativity in Preschool Age Children	Psychology of Aesthetics, Creativity and the Arts	N=75 children, aged 4 to 5, asked to complete a story stem and N=56 same age asked to draw a picture	Two authors and research assistant rated creativity on scale of 1-5	Cronbach's alpha .95 for stories, and .88 for the drawings
Pretz & Collum	2014	Self-Perceptions of Creativity Do Not Always Reflect Actual Creative Performance	Psychology of Aesthetics, Creativity and the Arts	N=90 4th year college undergrads. 3 creativity tasks: ideas for $1 million donation, photo caption, and essay on "dream" project	Ideas rated by 6 psych research assistants; "captions and essay also rated."	"Interraters reliabilities were" .862, .825, .892 and .731.
Baer, et al.	2010	Win or lose the battle for creative creativity: The power and perils of intergroup competition	Academy of Management Journal	N=280 undergrads, in 70 4 person groups, performed 2 idea generation tasks on improving student life	Ideas rated by 3 research assistants on creativity, scale 1-5	"the median interrater agreement coefficient (rwg); and two intraclass correlation coefficients: rwg2=.80; ICC=.37,

						ICC2=.64 ”
Silvia	2008	Discernment and Creativity: How Well Can People Identify Their Most Creative Ideas?	Psychology of Aesthetics, Creativity and the Arts	N=226 college psych students, created 4 divergent ideas, then chose the 2 most creative Total “expert” judges = 4 / 41	Ideas rated by undergrad research assistants on originality, scale 1-5	Not reported. Comparison was correlation between participants’ top 2 choices with judges ratings

REFERENCES

Akinci, C. & Sadler-Smith, E. 2012. Intuition in management research: A historical review. ***International Journal of Management Reviews***, 14(1): 104-122.

Amabile, T. M. 1982. Social psychology of creativity: A consensual assessment technique. ***Journal of Personality and Social Psychology***, 43(5): 997-1013.

Amabile, T. M. 1983. The social psychology of creativity: A componential conceptualization. ***Journal of Personality and Social Psychology***, 45(2): 357-376.

Amabile, T. M. 1996. ***Creativity in Context: Update to 'The Social Psychology of Creativity***. Boulder, CO: Westview Press.

Amabile, T. M. 2012. ***Componential Theory of Creativity*** to be published in Encyclopedia of Management Theory (Sage Pub. 2013, Kessler, E., Ed.), from working paper, 12-096, Harvard Business School, accessed June 2014 at http://www.hbs.edu/faculty/Pages/item.aspx?num=42469

Amabile, T. M., Hennessey, B., & Grossman, B. S. 1986. Social influences on creativity: The effects of contracted-for reward. ***Journal of Personality & Social Psychology*** 50(1): 14-23.

Amabile, T. M. & Khaire, M. 2008. Creativity and the role of the leader. ***Harvard Business Review***, 86: 100–109.

Andre, D. & Gobet, F. 2008. Sherlock Holmes—an expert's view of expertise. ***British Journal of Psychology***, 99:109-125.

Banks, M., Calvey, D., Owen, J. & Russell, D. 2002. Where the art is: Defining and managing creativity in new media SMEs. ***Creativity and Innovation Management***, 11(4): 255-264.

Baer, M. & Oldham, G. R. 2006. The curvilinear relation between experienced creative time pressure and creativity: Moderating effects of openness to experience and support for creativity. ***Journal of Applied Psychology***, 91(4): 963-970.

Barton, John H. 2003. Nonobviousness. ***IDEA*** 43(3): 475–506.

Batey, M. 2012. The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. ***Creativity Research Journal***, 24(1): 55-65.

Boden, M. A. 2004. ***The creative mind: Myths and mechanisms*** (2nd ed.). New York: Routledge.

Brogden, H.E., and Sprecher, T. B. 1964. Criteria of creativity, in C.W.Taylor (Ed.), ***Creativity, Progress and Potential***. New York: McGraw-Hill.

Bruner, J. S. 1962. The conditions of creativity. In H. E. Gruber, G. Terrell & M. Wertheimer (Eds.), ***Contemporary approaches to creative thinking: A symposium held at the University of Colorado*** (pp. 1-30). New York: Atherton Press.

Byron, K. & Khazanchi, S. 2012. Rewards and creative performance: A meta-analytic test of the theoretically derived hypotheses. ***Psychological Bulletin***, 138(4): 809-830.

Cader, R., Campbell, S. & Watson, D. 2005. Cognitive continuum theory in nursing decision-making. ***Journal of Advanced Nursing***, 49: 397-405.

Cameron, J. & Pierce, W. D. 1994. Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. ***Review of Educational Research***, 64(3): 363-423.

Chase, W. G. & Simon, H. A. 1973. Perception in chess. ***Cognitive Psychology***, 4: 55–81.

Conf. Board, 2008. Ready to Innovate, accessed July 2014 at https://www.conference-board.org/publications/publicationdetail.cfm?publicationid=1557.

Crandell, B., Klein, G., & Hoffman, R. R. 2006. ***Working minds: a practitioner's guide to cognitive task analysis*****.** London, U.K.: MIT Press.

Cropley, D. & Cropley, A. 2008. Elements of a Universal Aesthetic of Creativity. ***Psychology of Aesthetics, Creativity, and the Arts***, 2(3) 133-161.

Csikszentmihalyi, M. 1996. ***Creativity: Flow and the Psychology of Discovery and Invention***. New York, NY: Harper Collins.

Dailey, L. & Mumford, M. D. 2006. Evaluative aspects of creative thought: Errors in appraising the implication of new ideas. ***Creativity Research Journal***, 18: 367-384.

Dane, E. & Pratt, M. G. 2007. Exploring intuition and its role in managerial decision making. ***Academy of Management Review***, 32(1): 33-54.

Dane, E., Rockmann, K. W., and Pratt, M. G. 2012. When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. ***Organizational Behavior & Human Decision Processes***, 119(2): 187-194.

Denzin, N. K. & Lincoln, Y. S. (Eds.). 2011. ***Handbook of qualitative research*** (4th ed.). Thousand Oaks, CA: Sage.

Dhami, M. K. 2013. Judgment and Decision Making as a Skill: Learning, Development and Evolution. New York, NY: Cambridge University Press.

Dhami, M. K., Hertwig, R., & Hoffrage, U. 2004. The role of representative design in an ecological approach to cognition. ***Psychological Bulletin***, 130: 959-988.

Dhami, M. K. & Thompson, M. E. 2012. On the relevance of Cognitive Continuum Theory and quasirationality for understanding management judgment and decision making. ***European Management Journal***, 30: 316-326.

Doherty, M. E. & Kurz, E. M. 1996. Social judgment theory. ***Thinking and Reasoning***, vol. 2: 109-140.

Dorfler, v. & Ackermann, F. 2012. Understanding intuition: The case for two forms of intuition. ***Management Learning***, 43(5): 545-564.

Drury, J. L. & Darling, E. 2008. A "thin-slicing" approach to understanding cognitive challenges in real-time command and control. ***Journal of Battlefield Technology***, 11(1): 9-16.

Dunning, D., Johnson, K., Ehrlinger, J., Kruger, J. 2003. Why people fail to recognize their own incompetence. ***Current Directions in Psychological Science***, 12(3):83-87.

Dunwoody, P., Haarbauer, E., Mahan, R., Marino, C., & Tang, C. 2000. Cognitive adaptation and its consequences: A test of Cognitive Continuum Theory. ***Journal of Behavioral Decision Making***, 13: 35-54.

Eisenberger, R. & Aselage, J. 2009. Incremental effects of reward on experienced performance pressure: Positive outcomes for intrinsic interest and creativity. ***Journal of Organizational Behavior***, 30: 95-117.

Eisenberger, R. & Shanock, L. 2003. Rewards, intrinsic motivation, and creativity: A case study of conceptual and methodological isolation. ***Creativity Research Journal***, 15: 121-130.

Epstein, S. 2003. Cognitive experiential self-theory: An integrative theory of personality. In H. A. Tennen & J. A. Suls (Eds.), ***Handbook of psychology***. Hoboken, NJ: Wiley.

Epstein, S. 2010. Demystifying intuition: What it is, what it does, and how it does it. ***Psychological Inquiry***, 21(4): 295-312.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. 1996. Individual differences in intuitive-experiential and analytical- rational thinking styles. ***Journal of Personality and Social Psychology,*** 71: 390–405.

Ericsson, K. A. 2014. Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticisms. ***Intelligence***, 45: 81-103.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. 1993. The role of deliberate practice in the acquisition of expert performance. ***Psychological Review***, 100: 363–406.

Ericsson, K. A. & Ward, P. 2007. Capturing the naturally occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. ***Current Directions in Psychological Science***, 16: 346-350.

Ford, C. M. & Gioia, D. A. 2004. Factors influencing creativity in the domain of managerial decision-making. ***Journal of Management***, 26:705-732.

Florida, R.L. 2002. ***The rise of the creative class: And how it's transforming work, leisure, community, and everyday life***. New York: Basic Books.

George, J. & Zhou, J. 2007. Dual tuning in a supportive context: joint contributions of positive mood, negative mood, and supervisory behaviors to employee creativity. ***Academy of Management Journal***, 50: 605-622.

Gobet, F. & Simon, H. A. 1996. The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grand-master level chess. ***Psychological Science***, 7, 52–55.

Gore, J. & Riley, M. 2004. Recruitment and selection in hotels: Experiencing cognitive task analysis. In H. Montgomery, R. Lipschitz, & B. Brehmer (Eds.), ***How professionals make decisions***: 343-350. Mahwah, NJ: Lawrence Erlbaum.

Grabner, R. H. 2014. The role of intelligence for performance in the prototypical expertise domain of chess, ***Intelligence***, 45: 26-33.

Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. ***Psychological Review***, 108: 814-834.

Hamm, R. M. 1988. Moment-by-moment variation in experts' analytic and intuitive cognitive activity. ***IEEE Transactions on Systems, Man, & Cynernetics***, 18: 757-776.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. 1987. Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. ***IEEE Transactions on Systems, Man, and Cybernetics***, 17: 753-770.

Harvey, S. and Kou, C. 2013. Collective engagement in creative tasks: The role of evaluation in the creative process in groups. ***Administrative Science Quarterly***, 58(3): 346-386.

Hawking, S. & Mlodinow, L. (2010). ***The grand design.*** New York, NY: Harper & Row.

Henderson, J. M. & Hollingworth, A. 1999. High-level scene perception. ***Annual Review of Psychology***, 50: 243–271.

Hennessey, B. A. 2003. The Social Psychology of Creativity. ***Scandinavian Journal of Educational Research,*** 47(3): 253-271.

Hennessey, B. A. & Amabile, T. M. 2010. Creativity. ***Annual Review Psychology***, 61: 569-598.

Hoffman, R.R., Crandell, B., & Shadbolt, N. 1998. Use of critical decision method to elicit expert knowledge: A case study in the methodology of expert task analysis. ***Human Factors***, 40: 254-276.

Hoffman, R. R., Trafton, G., & Roebber, P. 2006. ***Minding the weather: How expert forecasters think***. Cambridge, MA: MIT Press.

Hoffman, R. R. & Woods, D. D. 2000. Studying cognitive systems in context. ***Human Factors***, 42(1): 1-7.

IBM, 2010. IBM global CEO study, accessed July 2014 at http://www-935.ibm.com/services/us/ceo/ceostudy2010/multimedia.html

Jackson. P.W. and Messick, S. 1965. The person, the product, and the response: Conceptual problems in the assessment of creativity. ***Journal of Personality***, 33(3): 309-329.

Kahneman, D. 2003. A perspective on judgment and choice: Mapping bounded rationality. ***American Psychologist***, 58(9): 697-720.

Kahneman, D. & Klein, G. 2009. Conditions for intuitive expertise: A failure to disagree. ***American Psychologist***, 64(6): 515-526.

Kahneman, D. 2011. ***Thinking, fast and slow***. New York, NY: Farrar, Strauss and Giroux.

Kaufman, J. C. & Baer, J. 2012. Beyond new and appropriate: Who decides what is creative? ***Creativity Research Journal***, 24(1): 83-91.

Kaufman, J. C., Baer, J., & Cole, J. C. 2009. Expertise, domains, and the consensual assessment technique. ***Journal of Creative Behavior***, 43: 223-233.

Kaufman, J. C., Baer, J., Cole, J., & Sexton, J. D. 2008. A comparison of expert and nonexpert raters using the consensual assessment technique. ***Creativity Research Journal***, 20: 171-178.

Kaufman, J. C., Baer, J., Cropley, D. H., Reiter-Palmon, R. & Sinnett, S. 2013. Furious activity vs. understanding: How much creative expertise is needed to evaluate creative work? ***Psychology of Aesthetics, Creativity, and the Arts***, 7(4): 332-240.

Klein, G. 1997. Naturalistic decision making: where are we now? In C.E. Zsambok, & G.A. Klein (Eds.). ***Naturalistic decision making***. Mahwah, NJ: Erlbaum.

Klein, G. A. 1998. ***Sources of power: How people make decisions.*** Cambridge, MA: MIT Press.

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

Klein, G.A., Calderwood, R., & Macgregor, D. 1989. Critical decision method for eliciting knowledge. ***IEEE Transactions on Systems, Man, and Cybernetics***, 19: 462-472.

Klein, G, Orasanu J., Calderwood R., Zsambok C. E. 1993. ***Decision making in action: Models and methods***. Westport, CT: Ablex.

Klein, H. & Myers, M. D. 1999. A set of principles for conducting and evaluating interpretive field studies in information systems. ***MIS Quarterly***, 23: 67-93.

Lincoln, Y. S. & Guba, E. G. 2000. Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), ***Handbook of qualitative research*** (2nd ed., pp. 163-188). Thousand Oaks, CA: Sage.

Litchfield, R. C. 2008. Brainstorming reconsidered: A goal-based view. ***Academy of Management Review***, 33: 649-668.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. 2001. Taking stock of naturalistic decision making. ***Journal of Behavioral Decision Making***, 14: 331-352.

Lonergan, D. C., Scott, G. M. & Mumford, M. D. 2004. Evaluative aspects of creative thought: Effects of appraisal and revision standards. ***Creativity Research Journal***, 16: 231-246.

Lubart, T. 2001. Models of the creative process: past, present and future. ***Creativity Research Journal***, 13: 295-308.

MacKinnon, D. W. 1978. ***In search of human effectiveness: Identifying and developing creativity***. Buffalo, NY: Creativity Education Foundation.

Madjar, N. & Oldham, G. R. 2002. Preliminary tasks and creative performance on a subsequent task: Effects of time on preliminary tasks and amount of information about the subsequent task. ***Creativity Research Journal***, 14(2): 239-251.

Mahan, R. P. 1994. Stress-induced strategy shifts toward intuitive cognition: A cognitive continuum framework approach. ***Human Performance***, 7(2): 85-118.

March, J. G. & Simon, H. A. 1993. ***Organizations***. Cambridge, MA: Blackwell.

Mathwicka, C., Malhotrab, N. K, & Rigdon, E. 2002. The effect of dynamic retail experiences on experiential perceptions of value: An internet and catalog comparison. ***Journal of Retailing***, 78: 51-60.

McAndrew, C. & Gore, J. 2007. "Convince me . . ." An interdisciplinary study of NDM and investment managers. ***Proceedings of the Eighth Conference on Naturalistic Decision Making***. Pacific Grove, CA.

McAndrew, C. & Gore, J. 2013. Understanding preferences in experience-based choice: a study of cognition in the "wild." ***Journal of Cognitive Engineering and Decision Making***, 7(2): 179-197.

Merriam, S. B. 2009. ***Qualitative research: a guide to design and implementation***. San Francisco, CA: Jossey-Bass

Miles, M. B. & Huberman, A. M. 1994. ***Qualitative data analysis: An expanded sourcebook***. Beverly Hills, CA: Sage

Militello, L. G. & Hutton, R. J. B. 1998. Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. ***Ergonomics***, 41(11): 1618-1641.

Militello, L. G., & Lim, L. 1995. Early assessment of NEC in premature infants. ***Journal of Perinatal and Neonatal Nursing***, 9: 1-11.

Mumford, M. D. 2003. Where have we been, where are we going? Taking stock in creativity research. ***Creativity Research Journal***, 15: 107–120.

Montag, T., Maertz, C. P., & Baer, M. 2012. A critical analysis of the workplace creativity criterion space. ***Journal of Management***, 38: 1362-1386.

Myers, M. D. 2009. ***Qualitative research in business and management***. Thousand Oaks, CA: Sage.

Neuman, W. L. 2005 ***Social research methods: Qualitative and quantitative approaches*** (4th ed.). Boston, MA: Allyn & Bacon.

Norris, P. & Epstein, S. 2011. An experiential thinking style: Its facets and relations with objective and subjective criterion measures. ***Journal Personality***, 75: 1043-1079.

Oldham, G. & Cummings, A. 1996. Employee creativity: personal and contextual factors at work. ***Academy of Management Journal***, 39: 607-634.

Paletz, S., and Schunn, C. 2010. A social-cognitive framework of multidisciplinary team innovation. ***Topics in Cognitive Science***, 2: 73–95.

Patel, V.L., & Groen, G.J. 1986. Knowledge-based solution strategies in medical reasoning. ***Cognitive Science***, 10: 91-116.

Patton, M. 1990. ***Qualitative Evaluation and Research Methods*** (2nd ed.). Thousand Oaks, CA: Sage

Pennington, N. & Hastie, R. 1993. A theory of explanation-based decision making. In G.A.

Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.). ***Decision making in action: models and methods***. Westport, CT: Ablex.

Prentky, P. A. 2000-2001. Mental illness and the roots of genius. ***Creativity Research Journal***, 13(1): 95-104.

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. 2001. Visual span in expert chess players: Evidence from eye movements. ***Psychological Science***, 12: 48–55.

Reiter-Palmon, R., Illies, M. Y., Cross, L. K., Buboltz, C., & Nimps, T. 2009. Creativity and domain specificity: The effect of task type of multiple index of creative problem-solving. ***Psychology of Aesthetics, Creativity and the Arts***, 3:73-80.

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. 2006. Productivity is not enough: A comparison of interactive and nominal brain-storming groups on idea generation and selection. ***Journal of Experimental and Social Psychology***, 42: 244–251.

Rhodes, M. 1961. An analysis of creativity. ***The Phi Delta Kappan***, 42(7): 305-310.

Roget, 2013. Creativity (n.d.). ***Roget's 21st Century Thesaurus, Third Edition*****.** Retrieved January 29, 2015 from Thesaurus.com website http://www.thesaurus.com/browse/creativity

Runco, M. A. 2004. Creativity. ***Annual Review Psychology***, 55: 657-687.

Runco, M. A. and Jaeger, G. J. 2012. The standard definition of creativity. ***Creativity Research Journal***, 24(1): 92-96.

Sadler-Smith, E. & Shefy, E. 2007. Developing intuitive awareness in management education. ***Academy of Management Learning and Education***, 6: 186-205.

Salas, E., Rosen, M. A., and Diaz-Granados, D. 2010. Expertise-Based Intuition and Decision Making in Organizations. ***Journal of Management***, 36: 941-973.

Santanen, E. L., Briggs, R. O. & Vreede, G. D. 2004. Causal relationships in creative problem solving: Comparing facilitation interventions for ideation. ***Journal of Management Information. Systems***, 20(4): 167-197.

Shalley, C. & Perry-Smith, J., 2001. Effects of Social-Psychological Factors on Creative Performance: The role of informational and controlling expected evaluation and modeling experience. ***Organizational Behavior & Human Decision Processes***, 84(1): 1-22.

Shalley, C. E., Zhou, J. and Oldham, G. R. 2004. The effects of personal and contextual characteristics on creativity: Where should we go from here? ***Journal of Management***, 30: 933-958.

Simonton, D. K. 2012. Taking the U.S. Patent Office criteria seriously: A quantitative three-criterion creativity definition and its implications. ***Creativity Research Journal***, 24: 97-106/

Singh, J. and Fleming, L. 2010. Lone inventors as sources of breakthroughs: Myth or reality? ***Management Science***, 56: 41–56.

Standing, M., 2008. Clinical judgment and decision-making in nursing—Nine modes of practice in a revised cognitive continuum. ***Journal of Advanced Nursing***, 62: 124-134.

Sternberg, R. J. & Lubart, T. L. 1999. ***The concept of creativity: Prospects and paradigms***. New York, NY: Cambridge University Press.

Sullivan, D. M. & Ford, C. M. 2010. The alignment of measures and constructs in organizational research: The case of testing measurement models of creativity. ***Journal of Business & Psychology***, 25: 505-521.

Taylor, C. W. 1988. Various approaches to and definitions of creativity, in: Sternberg, R. J. (Ed.), ***The Nature of Creativity: Contemporary Psychological Perspectives***. Cambridge University Press: MA.

Tversky, A. & Kahneman, D. 1974. Judgment under uncertainty: Hueristics and biases. ***Science***, 185: 1124-1131.

UK Department for Culture Media and Sport 2011. ***Creative Industries Economic Estimates***. Accessed August 2014 at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/77959/Creative-Industries-Economic-Estimates-Report-2011-update.pdf

US Bureau of Economic Analysis 2013. ***U.S. Bureau of Economic Analysis and National Endowment for the Arts Release Preliminary Report on Impact of Arts and Culture on U.S. Economy*** Accessed August 2014 at http://www.bea.gov/newsreleases/general/acpsa/acpsa1213.pdf

Van de Ven, A. H. 2007. ***Engaged scholarship: A guide for organizational and social research***. Oxford University Press.

West, M. A. 2002. Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups. ***Applied Psychology: An International Review***, 51(3): 355-424.

Yin, R. K. 2009. ***Case study research: Design and methods***. Thousand Oaks, CA: Sage.

Zhou, J. 2003. When the presence of creative coworkers is related to creativity: Role of supervisor close monitoring, developmental feedback, and creative personality. ***Journal of Applied Psychology***, 88(3): 413-422.

VITA

Michael Robert Seyle

Education

London School of Economics and Political Science, London, UK
Master of Science (M.Sc.), Cities – 2016-2019 (current)

Georgia State University, Robinson College of Business, Atlanta, GA
Executive Doctor of Business (EDB) – June 2015

San Diego State University, San Diego, CA
Masters of Business Administration (MBA) – May 2003

University of the Pacific, McGeorge School of Law, Sacramento, CA
Juris Doctor (JD, cum laude) – May 1989

Work Experience

President, US Region March 2017 to Present	BuroHappold Engineering New York, NY
President + CEO January 2009 to April 2017	Wimberly Allison Tong & Goo London, UK
Executive Vice President, CLO January 2000 to February 2004	Western Management Carlsbad, CA
Partner and Vice President September 1989 to December 1999	Seltzer Caplan McMahon Vitek San Diego, CA

Publication

Hunt, D. A. & Seyle, M. R. 1989. Carpenter v. United States: Securities Trading, Mail Fraud and Confidential Business Information—New Liability for Outsiders. 20 ***Pacific L. J.*** 839.

Research and Teaching Interests

Creativity, Innovation, Judgment and Decision-Making, Organizational Design and Behavior