Journal of the Association for Information Systems

Volume 7 | Issue 10 Article 30

11-6-2006

Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation

Douglas L. Dean
Brigham Young University, doug_dean@byu.edu

Jillian M. Hender
Henley Management College, Jill@hender.net

Thomas L. Rodgers
TRodgers@ieee.org

Eric L. Santanen Bucknell University
esantane@bucknell.edu

Recommended Citation
Dean, Douglas L.; Hender, Jillian M.; Rodgers, Thomas L.; and Santanen, Eric L. Bucknell University (2006) "Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation," *Journal of the Association for Information Systems*, 7(10), .
DOI: 10.17705/1jais.00106
Available at: https://aisel.aisnet.org/jais/vol7/iss10/30

Journal of the Association for Information Systems

Research Article

Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation [1]

Douglas L. Dean
Marriott School of Management
Kevin and Debra Rollins Center for e-Business
Brigham Young University
doug_dean@byu.edu

Jillian M. Hender
Henley Management College, UK
Jill@hender.net

Thomas L. Rodgers
Consultant
College Station, Texas
TRodgers@ieee.org

Eric L. Santanen
Management Department
Bucknell University
esantane@bucknell.edu

Abstract

Researchers and practitioners have an abiding interest in improving tools and methods to support idea generation. In studies that go beyond merely enumerating ideas, researchers typically select one or more of the following three constructs, which are often operationalized as the dependent variable(s): 1) idea quality, 2) idea novelty, which is sometimes referred to as rarity or unusualness, and 3) idea creativity. It has been chronically problematic to compare findings across studies because these evaluation constructs have been variously defined and the constructs have been sampled in different ways. For example, some researchers term an idea 'creative' if it is novel, while others consider an idea to be creative only if it is also applicable, effective, and implementable. This paper examines 90 studies on creativity and idea generation. Within the creativity studies considered here, the novelty of ideas was always measured, but in

[1] Detmar Straub was the accepting senior editor. This paper was submitted on September, 3, 2004, and went through three revisions.

some cases the ideas had to also meet additional requirements to be considered creative. Some studies that examined idea quality also assessed novelty, while others measured different quality attributes, such as effectiveness and implementability, instead. This paper describes a method for evaluating ideas with regard to four dimensions—novelty, workability, relevance, and specificity—and has identified two measurable sub-dimensions for each of the four main dimensions. An action-research approach was used to develop ordinal scales anchored by clearly differentiable descriptions for each sub-dimension. Confirmatory factor analysis revealed high loadings among the sub-dimensions that comprise each dimension as well as high discriminant validity between dimensions. Application of this method resulted in high inter-rater reliability even when the method was applied by different raters to different problems and to ideas produced by both manual methods and group support systems (GSS).

Keywords: idea evaluation, brainstorming, creativity, idea generation, group support systems, measurement

Introduction

For years, researchers and practitioners have studied methods of increasing the idea output of individuals and groups. Particular emphasis has been placed on improving the tools and methods used to support idea production because the ability to generate ideas is critical to promoting innovation and nurturing managerial problem-solving abilities. To this end, research has focused on manipulations that intend to increase both the quantity and the quality of ideas. According to Briggs et al. (1997), two challenges confront researchers wishing to evaluate the output of an idea generation process. First, a reliable way to rate each individual idea must be devised, which is especially difficult as in idea-generation studies the number of ideas commonly ranges from several hundred (e.g., Dennis et al., 1996) to more than a thousand (e.g., Hender et al., 2002). Second, the ratings of individual ideas must be aggregated into an overall score in order to assess the performance of the individual or group that produced the ideas.

Our review of the idea generation and creativity literature revealed that terms used to evaluate ideas can be grouped into the following four general constructs, as described by MacCrimmon and Wagner (1994): novelty, workability, relevance, and specificity. Different researchers have defined these terms variously, which has resulted in a proliferation of inconsistent definitions and related terms in the literature for each of these four constructs. Moreover, researchers sometimes sample the same construct in different ways. These irregularities have led to a number of problems. First, they make it difficult to train raters. Second, they can cause raters to be inconsistent in their individual ratings. Third, they can lead to inconsistencies between raters. Finally, they make it difficult to compare and generalize results across studies. This brings us to the three main questions addressed in this paper: first, what are appropriate component constructs for evaluating idea quality, novelty and creativity? Second, how can these constructs be assessed reliably? And, third, how should measures of individual ideas be aggregated to support comparisons across individuals or groups?

The first contribution of this paper is to distill constructs commonly used in the literature into a more manageable set. The second is to create and validate a reliable scale for each construct. The third is to recommend an approach for aggregating measures of individual ideas.

The next section of this paper provides additional theoretical background on the measurement of idea quality, novelty, and creativity. Then, we describe the development of our scales to measure dimensions of idea quality. Next, we evaluate our constructs and scales. Next, we suggest an implementation method for using our measures to summarize the output of an experiment. Then, we discuss our findings. Finally, we present our research limitations and conclusion.

Measures of Ideational Output

Early idea-generation research used quantity as a measure of quality, assuming that if a sufficient number of ideas were produced, the resulting idea pool would be more likely to contain high-quality ideas (Osborn, 1953). This positive correlation has been confirmed in some studies (Diehl and Stroebe, 1987; Gallupe et al., 1992; Valacich et al., 1993), but other research has found that the correlation between quantity and quality is tenuous (MacCrimmon and Wagner, 1994), and still others have posited that there is in fact a negative correlation between quantity and quality (Graham, 1977; Connolly et al., 1990; Gryskiewicz, 1980). Studies that go beyond merely enumerating ideas require researchers to select a definition of one or more of the three constructs that are typically operationalized as the dependent variable(s): 1) idea quality, 2) idea novelty, which is sometimes referred to as rarity or unusualness, and 3) idea creativity. However, as will be shown later in this paper, the definitions of these three constructs vary considerably among researchers. Therefore, we will now provide a succinct definition of each of these three constructs before examining each one in more detail.

Definition of Idea Quality

We define a quality idea as one that contains the following three characteristics. First, the idea should apply to the problem at hand (Aiken et al., 1996). Second, it should be an effective solution (Valacich, et al., 1995; Kramer and Kuo, 1997). Third, it should be implementable (Diehle and Stroebe, 1987). Each of these attributes is examined in detail later in this paper. According to this definition of idea quality, an idea can be termed a quality idea without it being novel or unusual, which is consistent with conventional definitions of a quality idea. In short, a quality idea is an implementable solution that will solve the problem, regardless of whether or not the idea itself is novel or unusual.

Definition of Idea Novelty

We define a novel idea as one that is rare, unusual, or uncommon (Connolly, Routhieaux, and Schneider, 1993). The most novel idea, then, is an idea that is totally unique; conversely, the least novel idea is the most common one (MacCrimmon and Wagner, 1994). In application, the novelty of any idea must be judged in relation to how uncommon it is in the mind of the idea rater or how uncommon it is in the overall population of ideas.

Definition of Idea Creativity

To define idea creativity, it is helpful to first examine the concept of creativity itself and to differentiate it from idea creativity. Creativity is typically viewed as a characteristic of an

environment, a process, a person, or a product (Rhodes, 1961). In terms of idea generation, environments, processes, persons, and groups that generate more novel ideas, or ideas that are not only novel but that also have other desirable attributes, are sometimes considered more creative than sources that produce fewer ideas with these qualities. Creativity can also be measured in terms of the characteristics of a product, such as an idea. Measures that apply to ideas, the product view, are the focus of this paper.

We define a creative idea as a quality idea that is also novel. That is, it applies to the problem, is an effective and implementable solution, and is also novel (MacCrimmon and Wagner, 1994).

Four General Dimensions of Creativity

Our definition of idea creativity is consistent with some researchers' definitions but is inconsistent with others. As shown in Table 1, definitions of creative products can be categorized into novelty-based definitions and multi-attribute definitions. In novelty-based definitions, ideas that are novel, sometimes referred to as original or rare, are considered creative regardless of whether they are applicable, effective, and implementable.

Conversely, in multi-attribute definitions of creativity, ideas must not only be novel; they must also have other quality attributes. For example, Plucker et al. (2004) suggest that creative products have two dimensions, namely, novelty and usefulness. They conducted a content analysis of 90 current articles, consisting of 10 each from business, education, and psychology journals, 30 from the Creativity Research Journal, and 30 from the Journal of Creative Behavior, all of which contained the term "creativity" in their title. These researchers found that, "*Overwhelmingly, the combination of novelty and usefulness were the most prevalent facets of both explicit and implicit definitions of creativity,*" and concluded that "*novelty and usefulness are two facets of creativity found in definitions both within our content analysis and when surveying the creativity literature in general*" (p.91).

Part 2 of Table 1 shows the relationships between various multi-attribute frameworks for creative ideas, proceeding from the least granular to the most granular in terms of specific attributes. As shown in Table 1, MacCrimmon and Wagner (1994) developed five dimensions of creative products. They based their analysis on research literature, practice literature, and U.S. Patent Office rules. The resulting framework is more specific than Plucker et al.'s framework. MacCrimmon and Wagner considered two facets of *originality*: *novelty,* which is defined as an idea that had not been previously expressed, and *non-obviousness*, an idea that was previously unknown even by people who are knowledgeable in the field. Their usefulness-related dimensions were also more specific than Pucker et al.'s, as they considered three additional sub-dimensions: *relevance,* the degree to which the idea actually applies to the specific problem domain; *workability*, the ability to implement the idea; and *thoroughness*, the extent to which ideas are fleshed out in detail in terms of being clear, concise, and exact. We chose the framework used by MacCrimmon and Wagner as the framework for this study, with the exception that our framework omits the non-obviousness sub-dimension because it was not used by any of the other studies in our literature review.

We had three specific reasons for choosing to adopt MacCrimmon and Wagner's dimensions as the framework for this study. First, their framework is more granular than that used by Plucker et al. Second, it is the most comprehensive, as it considers all of the constructs included in our definitions of idea quality, idea novelty, and idea creativity. Third, we agree with the fundamental assumptions about idea creativity that underlie this framework, mainly that mundane or common ideas are not creative. In effect, without novelty there is no creativity. Moreover, we hold that creative ideas must be useful to have practical value. Ideas that are not relevant to the problem, that are ineffective in solving the problem, or that cannot be implemented clearly cannot be usefully applied and so should not be considered creative ideas.

Table 1. Dimensions Identified in Creativity Literature

Study	Dimensions				
Part 1. Examples of Novelty-Centric Definitions of Creative Products					
Eisenberger and Selbst, 1994	Rarity				
Redmond et al., 1993	Originality				
Part 2. Examples of Multi-Attribute Definitions of Creative Products					
Woodman et al., 1993	Originality		Usefulness		
Plucker et al., 2004	Novelty		Usefulness		
Bessemer and Treffinger, 1981	Novelty		Resolution	Elaboration and Synthesis	
Wagner, 1996	Originality		Purpose	Implementation	
MacCrimmon and Wagner, 1994	Novelty	Non-obviousness	Relevance	Workability	Thoroughness
Current Study	Novelty		Relevance	Workability	Specificity

MacCrimmon and Wagner include the construct of thoroughness which, while not central to our definitions of idea quality or creativity, is certainly a desirable idea attribute because an idea that is expressed in specific and detailed terms is better than an idea that is expressed in vague terms. MacCrimmon and Wagner's definitions of these four dimensions are as follows:

- ***Novelty***: An idea is most novel if nobody has expressed it before.
- ***Workability***: An idea is workable if it does not violate known constraints or if it can be easily implemented.
- ***Relevance***: An idea is relevant if it satisfies the goals set by the problem solver.
- ***Thoroughness***: An idea is thorough if it is worked out in detail.

Since the concept of *thoroughness* may connote a solution that solves multiple aspects of the problem (which would be termed *effectiveness* in our framework), we have here renamed it "specificity" to avoid possible confusion.

The next section examines existing empirical studies in more detail in order to map their measures of idea quality, idea novelty, and idea creativity to these same fundamental dimensions.

Literature Review Sample Description

We examined articles that were published between 1990 and 2005. To be included in our literature review, articles had to meet two criteria: the studies had to be empirical and they must have counted ideas and/or subjectively evaluated idea quality or creativity in some way. To locate appropriate articles for consideration, we searched for the terms "brainstorming," "creativity," and "idea generation" under the topic of "management" in the ABI Inform Index as well as under the topic of "psychology" in the PsychInfo Index and in the Web of Science Index. Table 2 lists the ten journals with the greatest number of articles that met our criteria. We then searched each of these ten journals for all creativity and idea-generation articles from 1990 forward. This search located 79 articles that met our criteria. Finally, we also included 11 additional papers from other journals that met the criteria. These are also shown in Table 2. Therefore, the combined sample

Table 2. Journals and Articles included in Sample

#	Abbrev.	Journal	Articles
Journals With Most Occurrences of Idea-Generation Studies			
1	SGR	Small Group Research	16
2	I&M	Information & Management	13
3	OBHDP	Organizational Behavior and Human Decision Processes	9
4	JAP	Journal of Applied Psychology	8
5	JPSP	Journal of Personality and Social Psychology	8
6	JMIS	Journal of Management Information Systems	7
7	ISR	Information Systems Research	5
8	JCB	Journal of Creative Behavior	5
9	MISQ	Management Information Systems Quarterly	5
10	MS	Management Science	3
		Subtotal	79
Additional Journals Included in Sample			
11	AMJ	Academy of Management Journal	3
12	DS	Decision Sciences	2
13	ASQ	Administrative Science Quarterly	1
14	IEEESMC	IEEE Transactions Systems, Man Cybernetics	1
15	JM	Journal of Management	1
16	JMR	Journal of Marketing Research	1
17	OBHP	Organizational Behavior and Human Performance	1
18	PID	Personality and Individual Differences	1
		Subtotal	11
		Total	90

is composed of 90 journal articles in total. [2] The articles in the sample represent a cross-section of idea-generation and creativity studies and encompass ideas produced by verbal, written, and electronic idea-generation methods.

[2] The fact that our literature review resulted in the same number of articles as Plucker et al.'s (2004) is pure coincidence. While our literature review partially overlapped Plucker et al.'s, our

Most of the tasks considered in the articles are what Lamm and Trommsdorff (1973) term "means" tasks (i.e., subjects are asked to think of as many ways as possible to solve a given problem). Because idea quality and creativity cannot be determined objectively with these types of tasks, raters must assess the resulting ideas subjectively.

Ideation Evaluation Methods

We examined the methods described within the 90 articles to determine how idea quality, creativity, or both were measured. Table 3 shows the distribution of how ideation output was evaluated in the 90 articles.

Table 3. Articles using Ideation Evaluation Methods

Method	Abbrev.	Number	%
Ideas were counted but no quality or creativity measures were considered	CINQ	18	20
A single holistic measure of idea quality or creativity or separate holistic measures of both were considered	SHM	21	23
Quality or creativity dimensions were considered but not measured separately	CBNM	6	7
Quality or creativity dimensions were measured separately then combined explicitly	MSCE	11	12
Quality or creativity dimensions were reported separately	DRS	34	38
	Total	90	100

Counting Ideas

Of the 90 studies reviewed, 18 (20%) counted ideas but did not evaluate idea creativity or quality in any way (Crown and Rosse, 1995; Dennis and Valacich, 1993; Firestein, 1990; Gallupe et al., 1991; Gallupe et al., 1994; Gettys, 1987; Jessup et al., 1990; Jessup and Tansik, 1991; Kramer and Kuo, 1997; Leggett et al., 2000; MacCrimmon and Wagner, 1991-2; Offner et al., 1996; Paulus and Dzindolet, 1993; Paulus and Yang, 2000; Pinsonneault et al., 1999; Roy et al., 1996; Shepherd et al., 1995-6; Thornburg, 1991).

Single-Holistic Measure of Idea Evaluation

Twenty-one of the 90 articles used either a single measure of quality or a single measure of creativity. The authors of these articles did not discuss the sub-dimensions that were used to define creativity or quality or that needed to be met for an item to qualify as such. Five of the 21 articles used a single holistic measure to evaluate idea creativity (Basadur et al., 1982; Elam and Mead, 1990; Gaustello, 1995; Santanen et al., 2004; Shalley and Perry-Smith, 2001), while 15 of the 21 articles used a single holistic measure to evaluate idea quality (Alavi, 1993; Bouchard, 1972a; Dennis et al., 1999; Dennis et al., 1997; Dennis et al., 1996; Durand and VanHuss, 1992; Easton et al., 1990; Gallupe et al., 1988; Gallupe and McKeen, 1990; Marakas and Elam, 1997;

review included a much broader sample of journals. The fact that both literature reviews produced similar results in terms of emphasizing novelty and usefulness underscores how fundamental these constructs are in creativity research.

McGlynn et al., 2004; Nunamaker et al., 1991; Petrovic and Krickl, 1994; Smith, 1993; Valacich and Shearer, 1995; Wong and Aiken, 2003). Only Ocker et al. (1998) used both a single measure of quality and a separate single measure of creativity.

However, with holistic measures, raters may consciously or unconsciously include multiple constructs in a single rating. For example, one rater may intuitively include novelty or workability in his or her evaluation while another may not, an inconsistency that may lead to different ratings. Moreover, a single rater may be inconsistent across ideas because different constructs may seem more important to some ideas than to others. Thus, despite their efficiency, holistic measures do not address specific evaluation components in a predictable way.

Multidimensional Measures

We found that the four-dimension framework adopted by MacCrimmon and Wagner (1994) mapped well to the remaining 51 creativity and idea-generation articles. These articles all evaluated specific dimensions of ideas in some way. These 51 articles can be categorized as: 1) articles where dimensions of quality or creativity were considered but not measured, 2) articles where dimensions were measured separately and then combined explicitly, or 3) articles where dimensions were reported separately. These categories are presented here as evidence that the sub-dimensions considered by these 51 studies mapped to the four dimensions chosen in this research and as proof of the existence of different degrees of measurement explicitness.

Dimensions Considered but not Measured

Table 4 shows studies in which the dimensions were considered but not measured separately during the production of an overall rating. That is, raters were asked to think specifically about two or more dimensions while making a single aggregate rating that covered the multiple dimensions. This is an improvement over holistic measures because, although the dimensions were not measured separately, raters did consider specific subconstructs in some fashion. However, this method still has the disadvantage of not explicitly stating the relative importance of each dimension.

Dimensions Measured Separately and then Combined Explicitly

Some researchers have suggested that specific aspects of quality or creativity should be measured separately because they have found that some of these dimensions are not well correlated. For example, Rickards (1992) assessed the three dimensions of *novelty, feasibility*, and *potential*. A factor analysis showed that feasibility and potential were correlated but that neither construct was correlated with *novelty*. Similarly, Diehl and Strobe (1987) reported that originality and feasibility were not correlated, so they reported them separately. While past studies have shown that these dimensions are not well correlated, researchers sometimes combine them arithmetically for convenience when deriving a score for ideas. Table 5 shows studies in which multiple dimensions were measured but not reported separately. In these studies, separate measures were combined in some explicit way to produce an overall quality or creativity measure.

Table 4. Dimensions Considered but not Measured Separately (6 studies)					
#	**Study**	**Novelty**	**Workability**	**Relevance**	**Specificity**
1	Connolly et al., 1990	Creativity	Workability		
2	Eisenberger and Rhoades, 2001	Novelty		How well it dealt with problem	
3	Kramer and Kuo, 1997	Creativity	Feasibility	Effectiveness	
4 5	Shalley, 1991, Shalley, 1995	Novelty		Appropriateness	
6	Valacich, Wheeler, Mennecke and Wachter, 1995		Implementability	Ability to solve the problem	
	Total = 6	5	3	5	0

Table 5. Dimensions Measured Separately and then Combined Explicitly					
#	**Study**	**Novelty**	**Workability**	**Relevance**	**Specificity**
7	Barki and Pinsonneault, 2001	Originality	Feasibility	Effectiveness	
8	Diehl and Stroebe, 1991	Originality	Feasibility		
9	Mumford, 2001	Novelty, Unusualness	Logical, workability	Potential plausibility	
10	Gallupe et al., 1992	Originality	Feasibility		
11	Hender et al., 2002	Originality; Paradigm Relatedness			
12	MacCrimmon and Wagner, 1994	Novelty; Non-obviousness	Workability	Relevance	Thoroughness
13	Straus and McGrath, 1994		Feasibility	Impact	
14	Potter and Balthazard, 2004	Creativity	Feasibility	Relevance	
15	Sosik et al., 1997	Originality, Imaginativeness, Innovativeness		Value addition	
16	Cady and Valentine,1999	Novelty, Excitement	Adoptability; Non-violation of known constraints	Applicability, Ability to solve the problem, business potential	How well described
17	Miura and Hida, 2004	Novelty, Originality		Utility	
	Total = 11	10	8	8	2

Table 6. Dimensions Reported

#	Study	Novelty	Workability	Relevance	Specifity
18 19	Aiken and Vanjani, 1997, Aiken et al., 1996			Relation to topic	
20	Bouchard, 1972b		Practicality	Importance	
21	Cooper et al., 1998		Social acceptability		
22	Daily et al., 1996			Realistic	
23 24 25 26	Dennis, et al,. 1990; Dennis and Valacich, 1994; Valacich, Dennis, and Connolly, 1994; Valacich, George, Nunamaker, and Vogel, 1994			Magnitude of impact of policy on stakeholders	
27	Diehl and Stroebe, 1987	Originality	Feasibility		
28 29 30	Eisenberger and Selbst, 1994; Eisenberger and Armeli, 1997; Eisenberger, et al., 1998	Rarity			
31	Faure, 2004	Originality	Practicality	Effectiveness	
32	Fern, 1982			Goodness or usefulness for purpose	
33	Garfield et al., 2001	Novelty; Paradigm related			
34	Massetti, 1996	Novelty		Value	
35	Murthy and Kerr, 2003			Relevance	
36	Parent et al. 2000			Relevance; Usefulness	
37	Runco and Charles, 1993	Originality		Appropriatenes s	

Table 6. Dimensions Reported (Continued)					
#	**Study**	**Novelty**	**Workability**	**Relevance**	**Specifity**
38	Satzinger et al. 1999	Paradigm related			
39	Schweiger et al., 1986			Validity; Importance	
40	Shirani and Tafti, 1999				Solution based on facts and possibilities
41	Taylor et al., 1958		Feasibility; Probability	Effectiveness; Significance	Generality
42	Tung and Heminger, 1993			Validity and importance of assumptions	
43	Valacich et al., 1992			Importance	
44	Chirumbolo et al., 2005	Originality, Innovative			
45	Connolly, et al., 1993	Rarity			
46	Durand and VanHuss, 1992	Originality		Appropriateness	Detail, Depth, Clarity
47	Easton et al., 2003			Usefulness	
48	Kelly and Karau, 1993	Unusual, originality	Feasibility		
49	McLeod et al., 1996		Feasibility	Effectiveness	
50	Redmond et al., 1993	Originality		Quality	
51	Sosik et al., 1998	Originality			
	Total = 34	15	7	22	3

Compared to single holistic measures, this approach requires more assessments but has the advantage of combining components in a specific way. However, different combination methods have been used by different researchers, making it difficult to compare results across studies.

Dimensions Reported Separately

Table 6 shows studies in which either single or multiple dimensions were reported separately, an approach that allows for more transparency in the measurement of one or more specific dimensions.

Table 7 summarizes how frequently each dimension was considered as part of a holistic quality or creativity measurement (Table 4) or explicitly measured (Tables 5 and 6).

Table 7. Frequency of Dimensions Considered					
Table	Number of Studies	Studies that Considered or Sampled Each Dimension			
		Novelty	Workability	Relevance	Specificity
4	6	5	3	5	0
5	11	10	8	8	2
6	34	15	7	22	3
Total	51	31	18	35	5
Percentage	100%	59%	35%	69%	10%

Across these three tables, novelty and relevance were by far the most frequently used constructs, followed by workability; specificity was the least frequently used construct.

Conclusions from Literature Review

This examination of creativity and idea-generation literature yielded three conclusions. First, whether attempting to assess idea creativity or idea quality, different studies measured different constructs. Second, although a variety of constructs and methods have been used for idea assessment, these constructs map to one of the four primary dimensions identified by MacCrimmon and Wagner (1994): novelty, workability, relevance, and specificity. Finally, to systematically sample each dimension, it is better to score each dimension separately rather than assess a unitary, overall score. Thus, it is important to have a reliable way to measure each component dimension.

The following section describes the research approach used in this study to develop scales that assess each of these four dimensions.

Development and Application of Scales

Source of Ideas

The ideas used to develop the evaluation method described in this paper were gathered through two idea-generation experiments that we conducted simultaneously. One study, the results of which were reported in Hender et al. (2002), was undertaken with a group support system (GSS); the other was conducted in a manual environment. In the GSS environment, participants typed their ideas into the GSS, while in the manual environment, ideas were generated verbally and recorded on flip-chart paper by a facilitator. The pool of participants in the experiments was composed of 135 undergraduate students from a large American university. Participants were divided into 27 groups, each consisting of five students. Each group participated in both the GSS environment and the manual environment. Each group was assigned to use one of three techniques—brainstorming, analogies, or assumption reversals—and was given 20 minutes to generate ideas on each problem; each technique was used by a total of nine groups in each environment. Subjects were asked to generate as many ideas as possible and to be as creative as possible. The ideas that were generated in each treatment varied widely along the dimensions evaluated in this research.

In the GSS experiment, subjects generated ideas to solve the following problem: "A restaurant that is used by students is losing customers. What can the restaurant do to retain its customers?" In the manual experiment, subjects generated ideas to solve the following problem: "How can the city of Tucson attract more tourists?" The GSS experiment resulted in 2,105 ideas and the manual experiment resulted in 1,019 ideas. Prior to idea ratings, all non-ideas were removed, compound ideas were broken down into single ideas, and duplicate ideas were removed. In the end, 1,279 unique ideas remained for the restaurant problem and 692 unique ideas remained for the tourism problem, resulting in a total of 1,971 unique ideas.

Clarification of Dimensions and Sub-Dimensions

This section describes the approach used to create and refine the evaluation method. Two raters initially attempted to score a training sample of ideas based on MacCrimmon and Wagner's (1994) definitions of novelty, workability, relevance, and specificity, but were unable to achieve good inter-rater reliability using these definitions. It became evident that raters held different assumptions about what each dimension meant and that the raters were considering different aspects of each dimension. Trying to assess these different aspects with a single overall measure made measurement difficult and unreliable. Because this suggested the need to assess these aspects separately, we decomposed the initial four dimensions into more precise sub-dimensions, each sampling a different specific aspect of the four parent dimensions. To accomplish this, we re-examined the 51 studies in Tables 4, 5, and 6, and also consulted the creativity and idea-generation literature to see how the four general constructs had been decomposed by others, and we probed the assumptions held by our raters that had led to the variability in their ratings. This process produced the sub-dimensions described in the next sections.

Novelty

Novelty is a key construct for measuring the creativity of ideas. We now examine three different novelty-related constructs: rarity, originality, and paradigm relatedness.

Rarity. MacCrimmon and Wagner (1994) defined a novel idea as one that had not been previously expressed. According to this definition, then, a novel idea is unique or, at least, rare. The rarity of an idea can be determined by counting the number of times an idea occurs in a set of ideas. This approach, sometimes referred to as measuring the infrequency of an idea, measures the extent to which ideas are uncommon (Torrance, 1965; Lamm and Trommsdorff, 1973; Sosik et al., 1997; Sosik et al., 1998). One way to determine rarity is to count the number of subjects who proposed the same idea (Connolly et al., 1993). Thus, the rarer the idea, the lower its rarity score. Connolly et al. (1990) calculated rarity by computing the reciprocal of the number of idea occurrences. An advantage of this approach is that the rarer the idea, the higher its rarity score. A completely unique[3] idea has a rarity score of 1, with scores approaching zero for very

[3] The term *uniqueness* is used in two ways in the literature. First, it is used to describe an idea that is unique in the overall idea pool–that is, it is unique when considering the entire set of ideas generated by all groups in an experiment. Second, it is sometimes used to describe the non-redundant ideas that are produced by a single group (cf, (Parent, 2000; Connolly et al., 1993)). The term in the former sense is a measure of novelty for an idea relative to all ideas generated in the overall experiment. In the latter case, it refers to a non-redundant idea

common ideas. Others have counted an idea as being rare when it is produced by only one or two groups (Dennis, 1997) in an experiment.

The advantage of rarity-based approaches is that calculation is relatively simple, though similar ideas must be evaluated subjectively to determine whether or not they are repetitive. However, while measuring uniqueness or infrequency allows a researcher to see if many non-repetitive ideas are produced, this approach is limited in that an idea in an idea pool may be considered unique or rare even if it is in fact only slightly different from more common solutions to the problem. Because of this, rarity alone is insufficient as a measure of novelty.

Originality. To address the limitation of the rarity-based approach when used alone, we have defined original ideas as ideas that are not only rare but that also have the characteristic of being ingenious or imaginative. Thus, original ideas range from those that are common and mundane to those that are rare and imaginative. Thus, the most original ideas meet two criteria: First, they are rare. And, second, they are ingenious, imaginative, or surprising.

In terms of measuring originality, nine of the studies reported in Tables 4, 5, and 6 (summarized in Table 9) used rarity as a measure. In these tables, rarity is included as part of originality, even though it is a necessary but insufficient criterion of originality. Novelty or originality may also be assessed by raters using a rating scale, a method that was used in 20 of the studies.

Paradigm Relatedness. With reference to the creativity literature, several authors have deconstructed the dimension of novelty into sub-dimensions that reflect both originality and transformation power. For example, Besemer and Treffinger (1981) subdivide novelty into original (the product is unusual or infrequently seen) and transformational (the product is so revolutionary that it forces a shift in the way that reality is perceived). Jackson and Messick (1965) also state that the criteria for creative products include unusualness and transformation power. Similarly, Besemer and O'Quin (1987) break novelty down into two sub-dimensions: original (which includes descriptions such as novel, unusual, unique, and ingenious) and germinal (which includes descriptions such as trend setting, influential, revolutionary, and radical).

Nagasundaram and Bostrom (1994-95) and Gryskiewicz (1980) argue that originality is, by itself, an inadequate means of characterizing ideas. In addition to originality, they advocate considering *paradigm relatedness*, which is represented by the degree to which an idea relates to the currently prevailing paradigm, which is based on style of creativity (Kirton, 1976; 1987; 1989). Paradigm relatedness seems related to the concepts of *transformational* and *germinal*, as described above. Although Nagasundaram and Bostrom (1994-95) propose that paradigm relatedness is a construct orthogonal to originality, the two studies that have measured both constructs (Garfield et al., 2001 and Hender et al., 2002) have found that the two tend to move in concert.

---

produced by a single group. To avoid confusion, we recommend that the term *uniqueness* be reserved for cases in which ideas are unique to the entire idea pool and that the term *non-redundant* be used to refer to ideas produced by a single group. Authors should also make clear that non-redundant ideas are counted on a per group basis.

Nagasundaram and Bostrom (1994-95) developed a method of evaluating ideas that considered whether they were primarily adaptations (paradigm preserving (PP) ideas) or innovations (paradigm modifying (PM) ideas). Paradigm modifications in ideas may come about in two ways, each way acting individually or in concert, resulting in three different types of PM ideas (Figure 1). New ideas may be generated by introducing new elements into the problem context (quadrant 2), by altering the relationships between the elements of a problem (quadrant 3), or both (quadrant 4). Quadrant 4 ideas have the highest level of paradigm modifications, with quadrants 2 and 3 representing intermediate levels of paradigm modification.

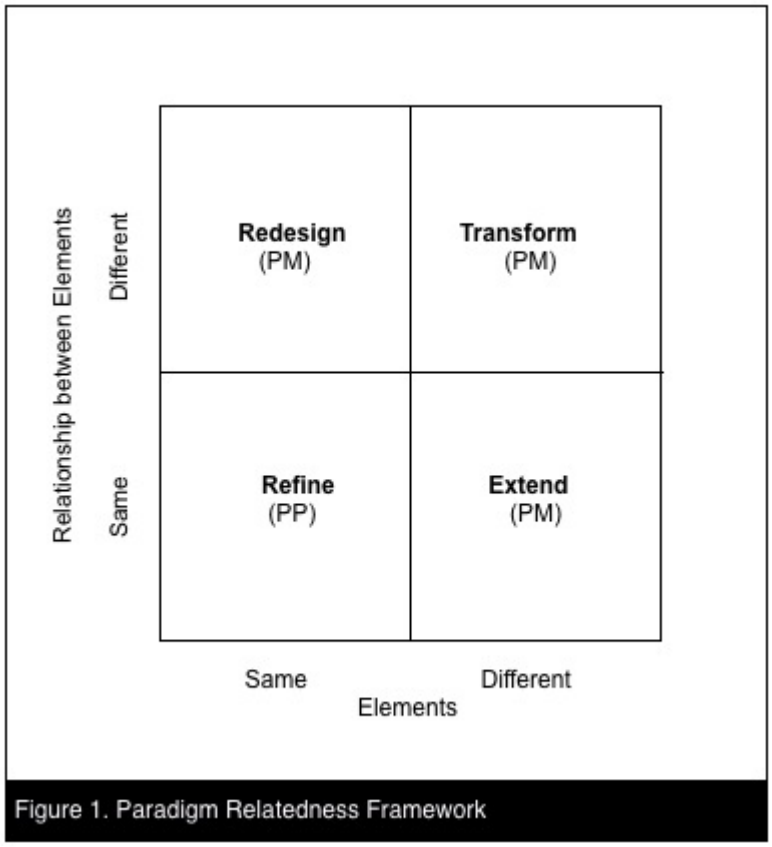

Figure 1. Paradigm Relatedness Framework

As an example of the application of the PP versus PM distinction, Satzinger et al. (1999) categorized ideas related to the question of how to solve a university parking problem as follows:

PP: solutions that deal with managing automobiles and the need to park them; for example, increasing the number of parking places.

PM: solutions that deal with transporting people to the university, approaches for delivering education to people away from the university, and solutions that focus on higher social goals.

While both originality and paradigm relatedness sample novelty, the two subconstructs have a somewhat different emphasis. The originality of ideas range from common and

mundane ideas to rare and imaginative ideas, whereas paradigm relatedness focuses on the degree to which an idea maintains or modifies a paradigm. Whether an idea maintains or modifies a paradigm is determined by whether new elements and relationships between elements are included in an idea.

Therefore, in accordance with Besemer and Treffinger (1981), Jackson and Messick (1965), and Besemer and O'Quin (1987), all of whom suggest that to be considered novel, an idea must be both original and transformational/germinal, we subdivide novelty into originality and paradigm relatedness.

Workability

According to MacCrimmon and Wagner (1994), the concept of workability is composed of two aspects: the idea is workable if it does not violate known constraints and if it can be easily implemented. These two aspects of workability have also been recognized by other authors. For example, in Briggs et al.'s (1997) study, ideas were evaluated according to their ease of implementation, and evaluators were instructed to consider economic, technical, and political feasibility. Diehle and Stroebe's (1987) definition of feasibility (preciseness and ease of implementation given available constraints) includes both aspects of feasibility/workability, which have also been recognized by other authors who have used this definition in other studies (Gallupe et al., 1992; Potter and Balthazard, 2004). Cady and Valentine (1999) define quality as, among other things, the degree to which an idea can be successfully adopted by an organization, where adoption incorporates the generation, development, and implementation of new ideas or behaviors. In addition, that idea must not violate known constraints. Other studies in Tables 4, 5, and 6 use one or the other of these aspects of workability. For example, Valacich et al. (1995) use implementability, while Cooper et al. (1998) use social acceptability. With reference to the creativity literature, Plucker et al. (2004) note that there is a social context related to the implementation of creative ideas. Creative ideas are initially considered novel but must ultimately become accepted in a particular cultural setting to have an impact.

Since an idea that cannot be implemented or that violates known constraints is not useful, workability is subdivided into implementability and acceptability.

Relevance

To be relevant, an idea must apply specifically to the problem at hand and it must be reasonable to expect that the idea will solve the problem. Researchers have previously used both applicability and effectiveness in relation to relevance. As shown in Tables 4, 5, and 6, some studies focus on applicability: its appropriateness (Shalley, 1991; Shalley, 1995), relation to topic (Aiken and Vanjani, 1997; Aiken et al., 1996), realism (Daily et al., 1996), usefulness for purpose (Fern, 1982), relevance (Murthy and Kerr, 2003; Parent et al., 2000), appropriateness (Runco and Charles, 1993), and validity (Schweiger et al., 1986). Others focus on effectiveness: how well the idea dealt with the problem (Eisenberger and Roades, 2001), its effectiveness (Kramer and Kuo, 1997; Barki and Pinsonneault, 2001; Faure, 2004), and its ability to solve the problem (Valacich et al., 1995). Some studies focus on both aspects simultaneously. For example, Cady and Valentine (1999) evaluated ideas based on their applicability as well as on their ability to solve the problem. With reference to the creativity literature, Besemer and Treffinger (1981) deconstructed their dimension of resolution into the sub-

dimensions of useful (the product has clear, practical applications) and valuable (the product is judged worthy by users, listeners, or viewers because it fills a financial, physical, social, or psychological need), as well as appropriate (the solution fits or applies to the problematic situation), and adequate (the product answers enough of the problematic situation).

Since an idea that is not expected to solve the problem, or that solves a different problem, is not useful in relation to the specific problem at hand, both applicability and effectiveness are important. Therefore, relevance is subdivided into applicability and effectiveness.

Specificity

Specificity refers to how well an idea is "thought out" and whether it has a complete, detailed, and elaborate description. MacCrimmon and Wagner (1994) developed this dimension from U.S. Patent Office specifications, which require that ideas be "full, clear, concise, and exact." Other researchers have emphasized different aspects of the specificity dimension. For example, Durand and VanHuss (1992) judged ideas on the basis of clarity, depth, and amount of detail, whilst Cady and Valentine (1999) judged them by how well they were described. With regard to the creativity literature, Besemer and O'Quin (1987) suggest that a creative product should be clear, complete, refined, and fluent. Besemer and Treffinger (1981) deconstructed their dimension of elaboration and synthesis into the sub-dimensions of expressive (the product is presented in a communicative, understandable manner), organic (the product has a sense of wholeness or completeness about it), and complex (the product or solution contains many elements at one or more levels).

Since ideas that are unclear, vague, incomplete, or that contain unclear causality, are less useful than ideas that are more specific in these areas, specificity is subdivided into clarity, completeness, and implicational explicitness. Table 8 contains our definitions of the general dimensions and sub-dimensions.

Mapping to Sub-Dimensions

Table 9 details the mapping of the constructs used in the 51 studies shown in Tables 4, 5, and 6 to the subconstructs operationalized in this study.

Table 9 demonstrates that studies have traditionally adopted a variety of approaches to idea evaluation. Originality (57%) and effectiveness (63%) were by far the most commonly sampled subconstructs, while applicability and workability subconstructs were sampled in only 39% and 33% of the articles, respectively. Specificity-related subconstructs were even less frequent and were sampled in only 10% of the studies, while paradigm relatedness was sampled in just 6% of the studies.

Both Type 1 and Type 2 articles focused on creativity, but while type 1 articles measured only novelty, Type 2 articles defined creativity as a combination of novelty and other quality attributes. Type 3 articles defined creativity and quality independently; that is, creativity included novelty only while quality did not include novelty at all. Type 4 articles focused on quality and defined quality as a combination of novelty and other quality constructs. The definition of quality in Type 5 articles did not include novelty but included more than one of the other quality dimensions. Finally, Type 6 articles defined quality in

terms of a single non-novelty quality dimension. Type 6 articles are different from the articles described under the single holistic measures in Section II. In Type 6 articles, quality is defined specifically as a single dimension instead of as an undefined measure that theoretically may or may not encompasses all possible quality dimensions.

Table 8. Our Definitions of the Quality Dimensions and Sub-dimensions

#	Dimension	Definition
1	***Novelty****	The degree to which an idea is original and modifies a paradigm.
1.1	Originality	The degree to which the idea is not only rare but is also ingenious, imaginative, or surprising
1.2	Paradigm relatedness	The degree to which an idea is paradigm preserving (PP) or paradigm modifying (PM). PM ideas are sometimes radical or transformational.
2	***Workability (Feasibility)***	An idea is workable (feasible) if it can be easily implemented and does not violate known constraints.
2.1	Acceptability	The degree to which the idea is socially, legally, or politically acceptable.
2.2	Implementability	The degree to which the idea can be easily implemented.
3	***Relevance****	The idea applies to the stated problem and will be effective at solving the problem.
3.1	Applicability	The degree to which the idea clearly applies to the stated problem.
3.2	Effectiveness	The degree to which the idea will solve the problem.
4	***Specificity***	An idea is specific if it is clear (worked out in detail).
4.1	Implicational explicitness	The degree to which there is a clear relationship between the recommended action and the expected outcome.
4.2	Completeness	The number of independent subcomponents into which the idea can be decomposed, and the breadth of coverage with regard to who, what, where, when, why, and how.
4.3	Clarity	The degree to which the idea is clearly communicated with regard to grammar and word usage.

Note: Our definitions of the novelty and relevance constructs are refinements to the definitions originally supplied by MacCrimmon and Wagner (1994).

Table 9. Article Characteristics Mapped to Studies

		Use of Terms				Constructs: Novelty (N)			Workability (W)		Relevance (R)		Specificity (S)			
#	Article Number in Tables 4, 5, & 6	Use "Creativity" term in title or motivation	Explicitly Define Creativity	Said Ideas are "Quality" or "Good"	Said Ideas are "Creative"	Rarity (1)	Originality (2)	Paradigm Related	Acceptability	Implementability	Applicability	Effectiveness	Implicational Explicitness	Completeness	Clarity	**Instructions-to-Subjects** (7)
Type 1. Creativity defined as novelty only																
1	28	■	■		■	■										N
2	29	■	■		■	■										N
3	30	■	■		■	■										N
4	38	■	■		■			■								nip
5	44	■	■		■		Rate 9									nip
6	51	■	■		■	■										N
7	11	■	■		■		Rate 4	■								N
8	33	■	■		■	■		■								V
9	45				(3)	■										V
Type 2. Creativity defined as novelty plus other constructs																
10	2	■	■		■		Rate 5					■				N
11	4	■	■		■		Rate 7				■	■				N, R
12	5	■	■		■		Rate 7				■	■				N
13	37	■	■		■	■					■	■				N, R
14	17	■			■		Rate10				■	■				N
15	34	■			■		Rate					■				N
16	48	■			■		Rate		■	■						N, W
17	12	■	■		■		Rate		■	■	■	■	■	■	■	N
18	9	■	■	■	(5)		Rate 5		■	■	■	■				V
Type 3. Define quality and creativity independently. Creativity defined as novelty only. Novelty is not included in quality																
19	46	■	■	■	■		Rate				■	■	■	■	■	nip
20	50	■	■	■	■		Rate					(6)				N
Type 4. Focus on quality . Quality defined as novelty plus other attributes																
21	14	■		(4)	■		Rate 5		■	■	■	■				tbsr
22	15	■		(4)	■	■						■				nip
23	1	■		■			Rate 7		■	■						N, W
24	3	■		■			Rate 7		■	■		■				tbsr
25	31	■		■		■			■	■		■				tbsr
26	7			■			Rate 7		■	■		■				tbsr
27	8			■			Rate 5		■	■						tbsr
28	10			■			Rate 5		■	■						tbsr
29	16			■			Rate10		■	■	■	■	■	■	■	N, W, R,S
30	27			■			Rate 5		■	■						bsr
Type 5. Focus on quality . Quality defined as multiple quality-related attributes. Novelty is not part of quality																
31	36			■							■	■				N, R
32	41	■		■					■	■	■	■	■	■	■	tbsr
33	49	■		■					■	■		■				tbsr
34	6			■						■		■				R
35	13			■					■	■		■				W, R
36	18			■							■					nip
37	19			■							■					N
38	20			■					■	■	■	■				tbsr
39	21			■					■							tbsr
40	22			■							■	■				R

Table 9. (Continued)

Type 6. Focus on quality. Quality defined in terms of a single non-novelty quality dimension																
41	23															R
42	24															R
43	25															V
44	26															R
45	32															nip
46	35															nip
47	39															R
48	42															nip
49	43															R
50	47															nip
51	40															nip
	Total	26	16	33	21	30		3	17	17	20	32	5	5	5	
	%	51%	31%	65%	41%	59%		6%	33%	33%	39%	63%	10%	10%	10%	

Table 9 Notes

1. In studies that evaluated rarity, a variety of different rarity thresholds were used.
2. The number of originality levels is shown when stated in the paper.
3. This article did not use the term "creativity" or claim that ideas were "creative;" however, since it measured only rarity, it was included in the creativity-centric articles.
4. Quality, defined as ideas that are novel and that also have one or more other quality attribute.
5. Quality, defined as ideas that are logical and feasible. Creativity defined as originality plus quality.
6. Quality was measured but not defined in the article.
7. Instructions-to-Subjects Legend
   - N = Subjects asked to be creative or to produce novel ideas
   - R = Subjects asked to produce relevant ideas
   - W = Subjects asked to produce workable ideas
   - T = Subjects asked to produce thorough ideas
   - tbsr = Traditional brainstorming rules
   - V = subjects asked to produce as many ideas as possible
   - nip = Instructions not in paper

The proportion of each type of article (Types 1 to 6) is shown in Table 10. In Table 10, in terms of creativity, an equal number of articles define creativity as novelty only (9) and as novelty plus other attributes (9). In articles that focus on quality attributes, articles that do not include novelty are more than twice as common (21) as those that do include novelty (10) (Chi-square = 3.903, p. = 0.048).

Table 10. Summary of Articles that Evaluate Idea Creativity or Quality

Type	1	2	3	4	5	6	Total
Number of articles	9	9	2	10	10	11	51
Percentage	18%	18%	4%	20%	20%	22%	100%

Of the 30 articles in Table 9 that assessed novelty in some way, significantly more used a rating approach than a rarity approach (21 and 9, respectively; Chi-square = 4.8, p = 0.028).

The right column in Table 9 indicates the type of instructions given to the participants in each study. Generally speaking, instructions that dealt with novelty were more frequent in creativity-focused studies, but this was not always the case. Sometimes instructions

that asked subjects to produce relevant or workable ideas corresponded to the measurement of these constructs. In addition, traditional brainstorming rules (Osborn, 1963) were sometimes used in both novelty-centric and non novelty-centric studies. In some papers, the authors did not describe the instructions. This is unfortunate because explicit descriptions of instructions would allow the manipulations and outcomes to be better understood by other researchers. Therefore, we recommend that future papers explicitly include this information.

Limitations in Comparability across Studies

As reflected in Table 9, there are a number of problematic limitations to comparing results across studies. First, studies that sample only novelty cannot determine whether or not the tested manipulations produced ideas that were workable, relevant, and specific. Conversely, since some studies examined only other constructs and not novelty, it is not known whether ideas produced from these manipulations were novel.

In addition, the problem of disparate construct sampling extends to non novelty-centric studies as well, with some sampling workability but not relevance and others sampling relevance but not workability. Regardless of their focus, comparability across studies would be improved if all of the studies sampled the same constructs.

Furthermore, novelty is evaluated in terms of rarity in some studies and in terms of originality (sometimes called creativity) in other studies. Moreover, studies that rate ideas for originality on ordinal scales use a variety of ranking levels, ranging from 4 to 10 levels. In addition, some rarity studies have indicators going in different directions—in some studies a high ranking indicates a rare idea, but in others a high ranking indicates a common idea.

Other factors also hinder comparability across studies, particularly the fact that the studies employed different experimental tasks, processes, and subjects. However, inconsistencies in constructs, subconstructs, and operationalizations are particularly vexing. Use of common constructs, subconstructs, and operationalizations would significantly reduce the difficulties involved in comparing across studies.

Development of Descriptive Anchors

Although refining the four primary dimensions into distinct sub-dimensions helped address the overloaded construct problem, there was still a need for a rating scheme for each sub-dimension that would support efficient and reliable ratings. Accordingly, we determined that it would be beneficial to have descriptive anchors to help raters more easily differentiate among ordinal levels. We developed a simple ordinal rating scheme for each sub-dimension so that each ordinal level would have a corresponding descriptive anchor and so each ordinal level would be clearly differentiated from other ordinal levels for the same construct. The number of ordinal levels included for each sub-dimension was determined by how well the descriptive anchors reflected separation among different levels. We struck a balance between differentiation and parsimony so that there would be an adequate number of levels to allow for clear distinctions and yet not so many as to be unmanageable or confusing.

We developed a four-point scale (1–4) for originality, paradigm relatedness, acceptability, implementability, applicability, and effectiveness, where a score of 1 denotes the lowest possible rating and a score of 4 indicates the highest possible rating. We developed a three-point scale (1–3) for completeness, and implicational explicitness and clarity. The choice of whether to create 3 or 4 ordinal levels for each construct was determined by the number of differentiated levels that naturally emerged during scale development.

We then combined the ratings from the sub-dimensions to create scales for each of the four main dimensions, as shown in Table 11.

Table 11. Derivations of General Constructs[4]

Construct	Range	Formula
Novelty	2-8	Originality + Paradigm relatedness
Workability	2-8	Acceptability + Implementability
Relevance	2-8	Applicability + Effectiveness
Specificity*	2-6	Completeness + Implicational explicitness

* As shown later in this paper, clarity was dropped from the specificity dimension based on our factor analysis.

We used the following procedure to develop clear, differentiating, descriptive anchors for the levels within each of the nine sub-dimensions:

1. We developed initial descriptive anchoring paragraphs using definitions of the sub-dimensions provided in the existing literature.
2. Two raters, working independently, scored a sample of ideas.
3. We calculated correlations using Pearson's correlation coefficient.
4. We identified ideas receiving significantly different scores. Low inter-rater reliability suggested the need to refine the descriptive paragraphs to achieve better differentiation, so, in these cases, raters examined the motivation behind their ratings and then further refined the descriptive paragraphs. The refined rating scheme was then applied to another sample of ideas.
5. We calculated scores for the four main dimensions.
6. We repeated the process until the descriptive paragraphs were adequately differentiated and the inter-rater reliability was greater than 0.7 for the four main dimensions. This process required approximately three iterations per dimension.

The descriptive anchors for the scoring scales are shown in Appendix A. With the exception of paradigm relatedness, the scoring scale definitions have the desirable characteristic of being independent of the problem used to generate ideas. That is, the rating scheme can be applied to ideas generated in relation to many different problems. The scoring scale definitions developed for paradigm relatedness were problem-specific, so different scales were applied to the restaurant problem and to the tourism problem used in this study, as shown in Appendix B.

Rating Ideas

[4] Calculating the means of ordinal levels and summing the ordinal levels of sub-constructs is common practice in behavioral sciences that consider situations, as in the current research, where ordinal scales approximate interval scales (Hand 1996, p. 463).

Once the evaluation method was refined, it was applied to the ideas produced by the GSS experiment. Two researchers rated these ideas, and then one of the original raters and a new rater evaluated the manually-generated ideas. The new rater required a few brief training rounds on a sample of ideas but was able to learn and reliably apply the rating scheme relatively quickly. During the rating of both idea sets, the specificity of the descriptive anchors helped the raters to be quicker and more methodical than they were without the proposed rating scheme.

The procedure for scoring the ideas was as follows: The same raters rated all dimensions for an idea set, and rating was spread out over a series of work sessions to minimize fatigue. All ideas were evaluated within one sub-dimension before rating began on a subsequent sub-dimension to help raters focus on one sub-dimension at a time, thereby avoiding potential across-dimension rating problems. Ratings given previously on other sub-dimensions were not visible during the rating of a specific sub-dimension. Ideas were presented to each rater one at a time on a computer screen that allowed the rater to enter a rating. Each rater independently scored the ideas using the descriptive anchors for each construct in Appendices A and B. Finally, the sub-dimension scores for each main dimension were combined to yield a main dimension score for each of the four main dimensions.

Evaluation of Constructs and Scales

Before we tested the overall model, we tested the sub-dimensions composing each hypothesized factor for reliability and construct validity. As shown in Table 12, the inter-rater reliability analysis using Chronbach Alpha resulted in good reliability between raters on each sub-dimension.

Table 12. Inter-rater Reliability on Sub-dimensions

Construct	Restaurant Ideas	Tourism Ideas
Originality	0.766	0.713
Paradigm relatedness	0.843	0.687
Acceptability	0.663	0.685
Implementability	0.713	0.714
Applicability	0.658	0.664
Effectiveness	0.729	0.667
Completeness	0.698	0.708
Implicational explicitness	0.783	0.690
Clarity	0.618	0.860

Correlation matrices for the eight items are presented in Tables 13 and 14. We highlight correlations between two items for each construct. Clarity did not correlate highly with completeness and implicational explicitness. For the remaining eight items—two for each construct—all correlations between items that measure the same construct are higher than all correlations between items that measure different constructs.

Table 13. Correlations Among Sub-dimensions, Restaurant Problem

Sub-dimensions	Originality	Paradigm relatedness	Accept-ability	Implem-entability	Applic-ability	Effective-ness	Clarity	Implicat-ional explic
Paradigm relatedness	0.563***							
Acceptability	-0.489***	-0.381***						
Implementability	-0.395***	-0.436***	0.472***					
Applicability	-0.051*	0.049*	0.197***	0.033*				
Effectiveness	-0.104***	-0.037*	0.275***	0.079***	0.539***			
Clarity	0.064***	-0.018	0.039*	0.006	0.012	0.010		
Implicational explicitness	0.183***	0.107***	-0.003	-0.010	0.371***	0.120***	0.032	
Completeness	0.265***	0.145***	-0.105***	-0.060***	0.225***	0.091***	0.041*	0.504***

*** p = 0.001; ** p = 0.01; * p = 0.05, N = 1,279
Note: Grey cells contain correlations between measures of related constructs. For example, originality and paradigm relatedness are both measures of novelty.

Table 14. Correlations Among Sub-dimensions, Tourism Problem

Sub-dimensions	Originality	Paradigm relatedness	Accept-ability	Implem-entability	Applic-ability	Effective-ness	Clarity	Implicat-ional explic
Paradigm relatedness	0.648***							
Acceptability	-0.346***	-0.424***						
Implementability	-0.37***	-0.498***	0.470***					
Applicability	-0.082*	-0.229***	0.332***	0.392***				
Effectiveness	-0.028	-0.095***	0.318***	0.190***	0.652***			
Clarity	0.157***	0.056*	0.039	-0.019***	0.199***	0.209***		
Implicational explicitness	0.170***	0.072*	-0.044	0.017***	0.118***	0.051*	0.039	
Completeness	0.276***	0.135***	-0.045*	0.030	0.176***	0.126***	0.196***	0.357***

*** p = 0.001; ** p = 0.01; * p = 0.05, N = 692

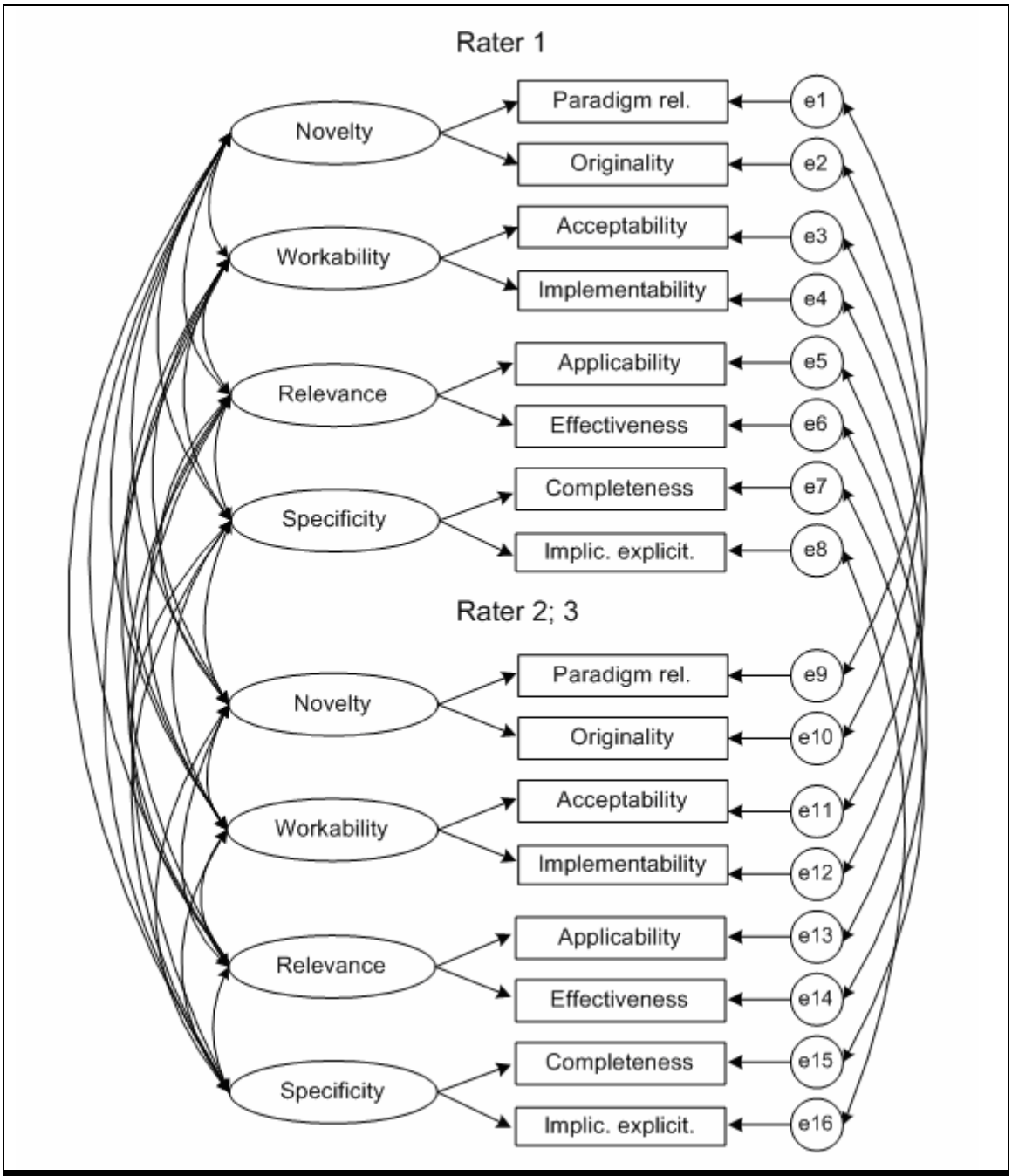

Figure 2. SEM Confirmatory Factor Analysis Model

Confirmatory Factor Analysis

We initially performed an exploratory factor analysis on the sub-dimensions for each set of ideas. The results were similar for both problems in terms of how sub-dimensions loaded onto factors. High loadings of sub-dimensions were related to each of the four main dimensions. Positive loadings ranged from 0.768 to 0.858; negative loadings ranged from -0.656 to -0.764. We extracted three factors using principle components analysis, and rotated them using the varimax rotation method. As expected, the applicability and effectiveness sub-dimensions loaded highly on one factor (relevance).

Also, completeness and implicational explicitness loaded strongly onto one factor (specificity), but clarity did not load strongly with specificity or any other factor, so we subsequently discarded it as a measure of idea quality. Both novelty items (originality and paradigm relatedness) and workability items (acceptability and implementability) loaded strongly onto the same factor, but while the novelty items loaded positively, the workability items loaded negatively. Consequently, we created separate factors for novelty and workability and tested the hypothesized model with confirmatory factor analysis.

We completed a confirmatory factor analysis of the model using structural equation modeling (SEM) performed with AMOS 5 and mPlus 3.11. AMOS calculates a much broader set of fit indices than mPlus, but mPlus explicitly supports use of ordinal observed variables. Analysis results from AMOS and mPlus were very similar, indicating that the use of ordinal data in AMOS did not significantly distort the model evaluation. [5] We present the results from AMOS in this paper because AMOS produces a more comprehensive set of fit indices. The structural model contains the sub-dimensions in Table 8, excluding clarity; therefore, it contains eight observed variables and four latent variables for the general dimensions. Figure 2 presents the SEM model used to conduct the confirmatory factor analysis. We tested the model separately for the restaurant and the tourism data sets.

Table 15. Standardized Maximum Likelihood Parameter Estimates of the Factor Loadings

Construct	Restaurant Ideas		Tourism Ideas	
	Rater 1	Rater 2	Rater 1	Rater 3
Novelty				
Originality	0.837	0.812	0.737	0.690
Paradigm relatedness	0.666	0.702	0.847	0.833
Workability				
Acceptability	0.742	0.764	0.610	0.777
Implementability	0.638	0.629	0.651	0.760
Relevance				
Applicability	0.999	0.916	0.969	0.986
Effectiveness	0.550	0.578	0.666	0.665
Specificity				
Completeness	0.598	0.641	0.694	0.699
Implicational explicitness	0.809	0.816	0.499	0.520

Note: We encountered a negative variance for rater 1 on the applicability observed variable in the restaurant data set. To manage this problem, we fixed the variance to a small positive value (0.001). This resulted in only minor changes in the fit variables (X^2 less than one) and produced a consistent estimate for the standardized parameter.

[5] For example, the Comparative Fit Index (CFI) calculated for the restaurant data was 0.97 from AMOS and 0.97 from mPlus and the CFI calculated for the tourism data was 0.96 from AMOS and 0.98 from mPlus.

Table 15 presents the factor loading parameter estimates for the measurement model. Overall, the findings indicate support for the hypothesized model. With a few exceptions, factor loadings are above 0.60. All loadings are highly significant (p = 0.001).

Model Fit Measures

As suggested by the literature (Byrne, 2001; Bentler and Bonett, 1980), we examined a variety of fit measures to determine the appropriateness of the models. The fit measures and suggested cut-off points for recognition of a well-fitting model are presented in Table 16.

Table 16. Model Fit Indices for Validity Testing

Item	Suggested Value	Restaurant (n = 1279)	Tourism (n = 692)
X^2		396.3	300.6
d.f.		69	68
X^2 significance (p-value)	$p > 0.05$	$p = 0.000$	$p = 0.000$
GFI (Goodness of Fit Index)$_1$	> 0.90	0.96	0.95
AGFI (Adjusted Goodness of Fit Index)$_1$	> 0.80	0.93	0.90
CFI (Comparative Fit Index)$_2$	> 0.90	0.97	0.96
NFI (Normed Fit Index)$_2$	> 0.90	0.97	0.95
RMSEA (Root Mean Square Residual)$_3$	< 0.08	0.06	0.07
RMR (Root Mean Square Error of Approximation)$_4$	< 0.10	0.03	0.02

Table 16 Fit Indices References: 1. (Joreskog and Sorbom, 1988); 2. (Bentler and Bonett, 1980); 3. (Browne and Cudeck, 1993); 4. (Hu and Bentler, 1995)

The results indicate strong support for the integrity of the model. The Chi-square statistic, which tests the hypothesis that the specification of the model is valid, is significant for both data sets. While the significant value of this statistic suggests that the fit of the data is not entirely adequate, a significant value of Chi-square is not unexpected since the statistic is a function of sample size.

The other fit statistics suggest an adequate fit of the model to the data. The root mean square residual (RMR), which represents the average residual value obtained from a comparison of the observed and predicted covariance matrices, is below 0.05 for both data sets. The goodness of fit index (GFI) is 0.95 or better for both data sets, and the adjusted goodness of fit index (AGFI) (adjusted for the number of degrees of freedom) is 0.90 or above for both data sets. These statistics compare the hypothesized model to a null model.

The range of the normed fit index (NFI) and the comparative fit index (CFI) is from 0 to 1.0. These indices compare the hypothesized model to a model in which all variables are independent of one another. The CFI adjusts for sample size. The values of the NFI and CFI fit statistics are well above 0.90 for both data sets. Finally, the root mean square error of approximation (RMSEA) is between 0.05 and 0.08 for both data sets, suggesting an acceptable fit to the data. This statistic indicates how well the model fits the population covariance model and takes into account the number of estimated model parameters. The inter-rater reliability for the main dimensions is shown in Table 17.

Table 17. Inter-rater Reliability on Main Dimensions (From SEM)		
Construct	**Restaurant Ideas**	**Tourism Ideas**
Novelty	0.927	0.891
Workability	0.910	0.831
Relevance	0.903	0.855
Specificity	0.985	0.886

Table 18. Inter-correlations among Dimensions				
Restaurant		Novelty	Workability	Relevance
rater 1	Workability	-0.771		
	Relevance	-0.036	0.156	
	Specificity	0.332	-0.090	0.510
rater 2	Workability	-0.811		
	Relevance	-0.012	0.283	
	Specificity	0.283	-0.016	0.398
Tourism		Novelty	Workability	Relevance
rater 1	Workability	-0.821		
	Relevance	-0.257	0.579	
	Specificity	0.262	-0.037	0.264
rater 3	Workability	-0.757		
	Relevance	-0.251	0.504	
	Specificity	0.151	0.025	0.235

Table 18 shows the correlations among the latent variables for the raters and data sets. In general, correlations among latent variables were low and consistent across raters; this reflects high discriminant validity among factors. This analysis also reflects a negative but high correlation between novelty and workability.

Ideas generated with manual support scored differently on the specificity dimension than ideas generated with the GSS. The average word length of ideas produced manually was 5.7 words; from the GSS environments the average word length was 12.3. Ideas produced in the manual environment averaged lower scores for implicational explicitness (1.10 vs. 1.46, $t_{1929} = 15.85$, $p = 0.000$), completeness (1.41 vs. 1.90, $t_{1790} = 17.71$, $p = 0.000$), and specificity (4.43 vs. 3.49, t1951 = 17.95, $p = 0.000$) than those produced with GSS. Manually-generated ideas were rated lower in implicational explicitness because the explanations of the ideas were generally too short to reflect implicational relations.

These results generally support our breakdown of the idea evaluation constructs for these two tasks. The factor loadings indicate that independent raters perceive differences among the four main dimensions, and these differences indicate good discriminant validity. Because this evaluation scheme was used successfully by different raters for two different problems and because use of the scheme produced consistent results, the reliability of the method is supported.

Recommended Application

The scales developed here can be used to guide future research in terms of constructs, rater training, and producing more comparable research findings. In Appendices A and B we have provided descriptions of ordinal levels and examples from both of our idea sets for each respective level. Appendix C provides suggestions for training raters. Finally, Appendix D provides guidance for idea counting conventions.

The remainder of this section provides recommendations regarding how to use consistent naming conventions and how to aggregate idea measures in a way that will improve comparability across studies. Since inconsistent naming conventions have produced the definitional inconsistencies described in our literature review, we recommend adopting the naming conventions depicted in Figure 3.

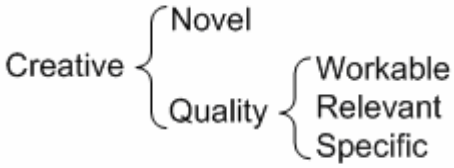

Figure 3. Relationships among Constructs

Specifically, we recommend that the term *quality* be used in studies that assess only non-novelty attributes. To avoid confusion between novelty-only studies and creativity studies where creativity is based on novelty plus other quality constructs, we recommend that the term *novelty*, and not *creativity*, be used for novelty-only studies. The term *creative* should be reserved for ideas that are novel and that also have other quality attributes[6]. Specificity is optional in terms of quality but should be measured when specificity is included in the focus of a particular study.

Table 19 summarizes the measures that we recommend be adopted in the three types of studies described in this paper: quality, novelty-only, and creativity. We recommend that authors of each type of study report on all the indicated measures, not just on the ones that show significant differences. This will increase the comparability of findings across studies.

The measures in Table 19 are divided into three types: counts, summation measures, and averages.

Counts. When counting the number of ideas produced by groups, only non-redundant ideas should be counted for each group to avoid counting the same idea multiple times. The terms *quality* ideas, *novel* ideas, and *creative* ideas should be reserved for ideas

[6] We recognize that some past authors have chosen to use the term *good*, or sometimes *quality*, for what we call *creative*. We have chosen this term because it is consistent with the findings of our literature review, because it is internally consistent with our naming convention, and because of the semantic ambiguity inherent in the term *good*.

Table 19. Recommended Measures

Measure	Focus of Study		
	Quality	Novelty Only	Creativity
Counts			
Ideas	X	X	X
Quality Ideas	X		
Novel Ideas		X	X
Creative Ideas			X
Summation Measures			
Total Quality	X		
Total Novelty		X	
Total Creativity			X
Averages			
Average Quality	X		
Average Novelty		X	X
Average Creativity			X

Table 20. Percent of Ideas Conforming to Each Level

Threshold	Quality		Novel		Creative	
	% ideas with this score	Cumu-lative %	% ideas with this score	Cumul-ative %	% ideas with this score	Cumul-ative %
Restaurant Ideas; Generated with GSS; N = 1279						
8	0.5%	0.5%	4.8%	4.8%	0.0%	0.0%
7	12.4%	13.0%	4.6%	9.4%	0.5%	0.5%
6	46.9%	59.9%	14.5%	23.8%	8.7%	9.1%
5	26.3%	86.2%	34.5%	58.3%	39.1%	48.2%
4	7.8%	94.1%	25.4%	83.7%	30.7%	79.0%
3	2.8%	96.9%	10.3%	94.1%	12.1%	91.1%
2	3.1%	100%	5.9%	100%	8.9%	100%
Tourism Ideas; Manually Generated; N = 692						
8	2.3%	2.3%	4.9%	4.9%	0.0%	0.0%
7	12.3%	14.6%	2.9%	7.8%	0.4%	0.4%
6	40.9%	55.5%	15.9%	23.7%	9.1%	9.1%
5	22.1%	77.6%	27.7%	51.4%	28.6%	37.7%
4	16.3%	93.9%	29.8%	81.2%	37.7%	75.4%
3	2.3%	96.2%	13.3%	94.5%	15.5%	90.9%
2	3.8%	100.0%	5.5%	100%	9.1%	100%

that meet specific thresholds of the constructs included in these measures. To identify novel ideas, quality ideas, and creative ideas, we recommend a threshold approach similar to that introduced by Diehl and Stroebe (1987) [7] and later used by Dennis et al. (1997). That is, quality ideas should meet a specific threshold on workability and relevance. Creative ideas should meet a specific threshold on novelty, workability, and

[7] Diehl and Stoebe categorized an idea as *good* if it received their highest rating for originality (part of novelty) and at least the penultimate rating for feasibility (part of workability).

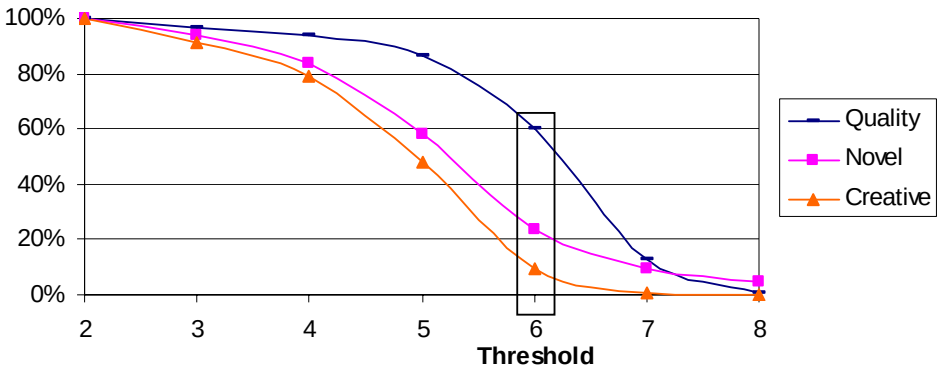

Figure 4. Percent of ideas meeting each threshold (restaurant ideas)

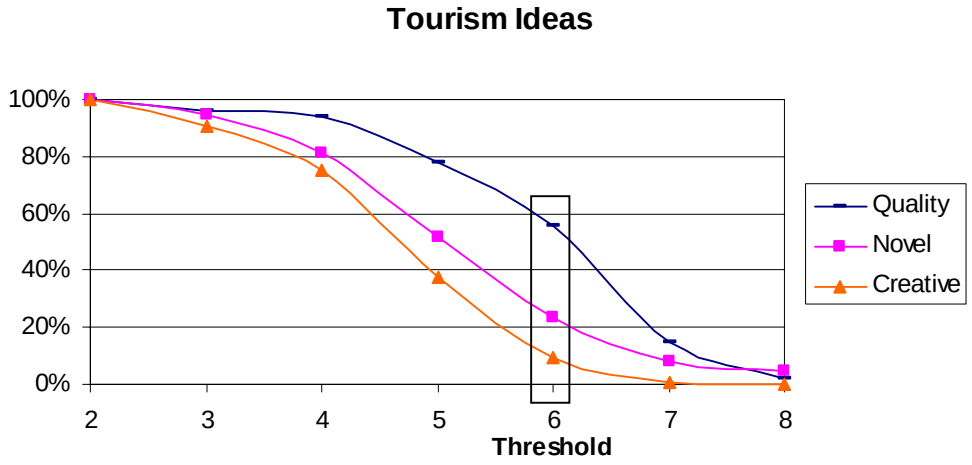

Figure 5. Percent of ideas meeting each threshold (tourism ideas)

relevance. Threshold measures are noncompensatory measures. That is, strength in one quality indicator cannot compensate for weakness in another area.

To identify a reasonable threshold, we performed a sensitivity analysis to determine the percentage of ideas that would qualify as having quality, being novel, and being creative when required to meet specific measurement thresholds in our two idea samples. The results are shown in Table 20 and in Figures 4 and 5.

For novelty, if a threshold of eight were to be used, it would mean that only one in 20 ideas would be novel. Likewise, a threshold of seven would mean that only one in about 11 ideas would be novel. A threshold of six means that ideas are pretty novel but they don't have to rate at the top on at least one of the novelty scales (a four for originality or a four for paradigm relatedness) to be considered novel. A threshold of six means a score of three out of four on both originality and paradigm relatedness would qualify an

idea as novel. This seems reasonable. In our data sets this would mean that about one in five ideas qualify as novel.

For creativity, no ideas met a threshold of eight for all three constructs. Only about 0.5% met a threshold of seven for the three constructs. These thresholds are obviously too restrictive. Conversely, a threshold of five means that 48.2% of restaurant ideas and 37.7% of tourism ideas would be considered creative. This is logically inconsistent with the requirement that a creative idea must also be novel, where novelty includes rarity by definition. By requiring a threshold of six, creativity will include only novel terms as per the novelty threshold. Thus, a threshold of six seems an appropriate compromise. In our datasets this threshold resulted in 9.1% of ideas qualifying as creative.

We acknowledge that methods are sometimes used to create novel ideas before these ideas are made workable and relevant. The problem is that a measurement that considers only novelty cannot determine the extent to which the ideas are workable and relevant. For example, one method may produce novel ideas that meet other quality criteria while another method may not. More complete measurement and reporting will make this distinction more visible and will improve the accuracy of comparability across studies.

Summation Measures. Summation measures sum the novelty, quality, or creativity of all ideas. First, for each idea, the novelty, quality, or creativity is determined. Then this value for all ideas is summed. When multiple measures, such as the idea's novelty, workability, and relevance scores, are assessed for a single idea, these can be averaged, to obtain a creativity score. Summation measures are compensatory measures in that strength in one construct can compensate for a weakness in another construct. Also, summation measures are blended measures in that they take into account both the quantity of ideas and the value of all ideas.

Averages. Averages of idea quality, novelty, and creativity should also be calculated and reported. Average specificity, though optional, can be an important indicator when ideas are produced through methods that facilitate or constrain communication, as in the case reported in this study in which we compare the specificity of ideas generated in a manual versus a GSS environment. In addition, the clarification of specificity constructs provided in this study can be used to promote the generation of clearly stated and well-developed ideas.

Discussion

The measurement approach proposed here is more comprehensive than those used in other studies. In addition, although the eight subconstructs are more granular and concisely defined than those used in many previous studies, the number of ordinal levels in each subconstruct remains small and manageable (3 or 4). This, combined with the fact that ordinal levels are given descriptive anchors, will help in the training of new raters.

Measuring all three general constructs allows ideation studies to assess not only novelty, but also whether ideas were produced that were workable and relevant. This study provides a novelty rating approach that includes rarity. This avoids the problem of giving equal credit for incremental but rare ideas versus more radical and rare ideas.

The term creativity has been used in many different ways in past studies, falling generally into the novelty-centric and quality-centric definitions. This study shows that creativity is sometimes described as novelty and at other times is defined as novelty plus other quality attributes. Our approach resolves this definitional inconsistency by providing a portfolio of measures that are serviceable for both perspectives. Adoption of this approach will allow findings to be compared across both types of studies.

In our research, higher values always correspond to higher levels of a construct in our scales. Adoption of this and of using a consistent number of ordinal levels in future studies would eliminate some of the problems that are encountered when comparing studies that rate constructs differently and that do not use the same number of ordinal levels. Adopting the definitions and rating scales proposed here should increase comparability across future studies in this important research area.

Limitations and Future Research

We used only two raters to evaluate ideas from each experiment, an approach that is consistent with many other studies (e.g., Dennis et al., 1999; Santanen et al., 2000; Satzinger et al., 1999). Future studies could examine our assessment approach with additional raters. While we have attempted to provide helpful guidelines, the idea-counting protocol and rating scales provided by this research still require an amount of subjective assessment. Different raters may have different worldviews, biases, and assumptions. Therefore, careful training and inter-rater refinement, as described in Appendix C, are very important. Future studies could also apply our approach to ideas produced for other kinds of problems.

Conclusion

In past studies, a variety of measures have been used to evaluate ideas, and each measure had its own set of limitations. In particular, the definitional inconsistency among the definitions and operationalizations of these measures has led to great difficulty in correctly interpreting previous studies; this, in turn, will result in inconsistencies in future studies. This study has attempted to address some of these limitations. A multi-dimensional measure has been proposed that would potentially support future studies by allowing researchers to select some or all of the sub-dimensions, as appropriate. Furthermore, clear definitions that are consistent with these operationalizations are provided. The metrics provided by this study can help researchers in two ways. First, the measures sample a balanced set of dimensions in a predictable, reliable way. A combination of these measures supports meaningful aggregate measures of ideational output. Adoption of the methods described in this paper would improve the comparability of findings across future studies, making it easier to compare the effects of different idea-generation manipulations. Second, the anchored scales should make it easier to train raters to rate ideas reliably and to facilitate the training of new raters. This contribution is important, especially given the effort that has previously been necessary to train new raters. An understanding of these constructs, and of how to measure them, can facilitate future research and management of the evaluation of ideas.

Acknowledgements

The authors express appreciation to Paul Paulus, Scott Poole, Beth Hennessey, Robert Briggs, and Christopher Barlow who provided helpful suggestions to improve this paper. The authors also thank James Anderson and Joseph Olsen for statistics consulting. The authors received funding support from the Marriott School of Management at Brigham Young University and a grant from the Economic and Social Research Council, UK.

References

Aiken, M. and M. Vanjani (1997) "A Comparison of Synchronous and Virtual Legislative Session Groups Faced with an Idea Generation Task," *Information & Management* (33) 1, pp. 25-31.

Aiken, M., M. Vanjani, and J. Paolillo (1996) "A Comparison of Two Electronic Idea Generation Techniques," *Information & Management* (30) 2, pp. 91-99.

Alavi, M. (1993) "An Assessment of Electronic Meeting Systems in a Corporate Setting," *Information & Management* (25) 4, pp. 175-182.

Barki, H. and A. Pinsonneault (2001) "Small Group Brainstorming: Is Electronic Brainstorming the Most Effective Approach?" *Small Group Research* (32) 2, pp. 158-205.

Basadur, M., G. B. Graen, and S. G. Green (1982) "Training in Creative Problem Solving: Effects on Ideation and Problem Finding and Solving in an Industrial Research Organization," *Organizational Behavior and Human Performance* (30) 1, pp. 41-70.

Bentler, P. and D. Bonnett (1980) "Significance Test and Goodness of Fit in the Analysis of Covariance Structures," *Psychological Bulletin* (88) 3, pp. 588-606.

Bessemer, S. P. and K. O'Quin (1987) Creative Product Analysis: Testing a Model by Developing a Judging Instrument" in S. G. Isaksen (ed.) *Frontiers of Creativity: Beyond the Basics*, Buffalo, New York: Bearly.

Bessemer, S. P. and D. J. Treffinger (1981) "Analysis of Creative Products: Review and Synthesis," *Journal of Creative Behavior* (15) 3, pp. 158-178.

Bouchard, T.J., and Hare, M. (1970). "Size, Performance, and Potential in Brainstorming Groups," *Journal of Applied Psychology*, (54) 1, pp. 51-55.

Bouchard, T. J. (1972a) "A Comparison of Two Group Brainstorming Procedures," *Journal of Applied Psychology* (56) 4, pp. 418-421.

Bouchard, T. J. (1972b), "Training, Motivation, and Personality as Determinants of the Effectiveness of Brainstorming Groups and Individuals," *Journal of Applied Psychology,* (56), pp. 324-331.

Briggs, R., B. Reinig, M. Shepherd, J. Yen et al. (1997) Quality as a Function of Quantity in Electronic Brainstorming. *Proceedings of Thirtieth Hawaii International Conference on System Sciences, 1997*, pp. 94-103 2.

Browne, M. W. and R. Cudeck (1993) "Alternative Ways of Assessing Model Fit" in K. A. Bollen and J. S. Long (Eds.) *Testing Structural Equation Models*, Newbury Park, CA: Sage, pp. 445-455.

Byrne, B.M. (2001), *Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming*, Lawrence Erlbaum Associates, London.

Cady, S. H. and Valentine, J. (1999), "Team Innovations and Perceptions of Consideration," *Small Group Research*, (30) 6, pp. 730-750.

Campbell, J.P. (1968), "Individual Versus Group Problem Solving in an Industrial Sample," *Journal of Applied Psychology* (52) 3, pp. 61-91.

Chirumbolo, A., Mannetti, L., Pierro, A., Areni, A. and Kruglanski, A. W. (2005), "Motivated Closed-Mindedness and Creativity in Small Groups," *Small Group Research*, (36) 1, pp. 59-82.

Connolly, T., L. M. Jessup, and J. S. Valacich (1990) "Effects of Anonymity and Evaluative Tone on Idea Generation in Computer-Mediated Groups," *Management Science* (36) 6, pp. 689-703.

Connolly, T., Routhieaux, R. L. and Schneider, S. K. (1993), "On the Effectiveness of Group Brainstorming: Test of One Underlying Cognitive Mechanism," *Small Group Research*, (24) 4, pp. 490-503.

Cooper, W. H., R. B. Gallupe, S. Pollard, and J. Cadsby (1998) "Some Liberating Effects of Anonymous Electronic Brainstorming," *Small Group Research* (29) 2, pp. 147-178.

Crown, D. F. and Rosse, J. G. (1995), "Yours, Mine and Ours: Facilitating Group Productivity Through the Integration of Individual and Group Goals," *Organizational Behavior and Human Decision Processes*, (64) 2, pp. 138-150.

Daily, B., A. Whatley, S. R. Ash, and R. L. Steiner (1996) "The Effects of a Group Decision Support System on Culturally Diverse and Culturally Homogeneous Group Decision Making," *Information & Management* (30) 6, pp. 281-289.

Dennis, A. R., J. E. Aronson, W. G. Heninger, and E. D. W. II (1999) "Structuring Time and Task in Electronic Brainstorming," *MIS Quarterly* (23) 1, pp. 95-108.

Dennis, A. R., and J. S. Valacich (1993), "Computer Brainstorms: More Heads Are Better Than One?" *Journal of Applied Psychology,* (78) 4, pp. 531-537.

Dennis, A. R. and J. S. Valacich (1994) "Group, Sub-Group, and Nominal Group Idea Generation: New Rules for a New Media?" *Journal of Management* (20) 4, pp. 723-736.

Dennis, A. R., J. S. Valacich, T. Carte, M. Garfield et al. (1997) "Research Report: The Effectiveness of Multiple Dialogues in Electronic Brainstorming," *Information Systems Research* (8) 2, pp. 203-211.

Dennis, A. R., J. S. Valacich, T. Connolly, and B. E. Wynne (1996) "Process Structuring in Electronic Brainstorming," *Information Systems Research* (7) 2, pp. 268-277.

Dennis, A. R., J. S. Valacich, and J. F. Nunamaker, Jr. (1990) "An Experimental Investigation of Small, Medium and Large Groups in an Electronic Meeting System Environment," *IEEE Transaction Systems, Man Cybernetics* (25) 5, pp. 1049-1057.

Diehl, M. and W. Stroebe (1987) "Productivity Loss in Brainstorming Groups: Toward the Solution of a Riddle," *Journal of Personality and Social Psychology* (53) 3, pp. 497-509.

Diehl, M. and Stroebe, W. (1991), "Productivity Loss in Idea-Generating Groups: Tracking Down the Blocking Effect," *Journal of Personality and Social Psychology*, (61) 3, pp. 392-403.

Durand, D. E. and S. H. VanHuss (1992) "Creativity Software and DSS," *Information & Management* (23) 1, pp. 1-6.

Easton, G., Easton, A. and Belch, M. (2003), "An Experimental Investigation of Electronic Focus Groups," *Information & Management*, (40) 8, pp. 717-727.

Eisenberger, R. and Selbst, M. (1994), "Does Reward Increase or Decrease Creativity?" *Journal of Personality and Social Psychology*, (66) 6, pp. 1116-1127.

Eisenberger, R. and Armeli, S. (1997), "Can Salient Reward Increase Creative Performance Without Reducing Intrinsic Creative Interest?" *Journal of Personality and Social Psychology*, (72) 3, pp. 652-663.

Eisenberger, R., Armeli, S. and Pretz, J. (1998), "Can the Promise of Reward Increase Creativity?" *Journal of Personality and Social Psychology*, (74) 3, pp. 704-714.

Eisenberger, R. and Rhoades, L. (2001), “Incremental Effects of Reward on Creativity,” *Journal of Personality and Social Psychology*, (81) 4, pp. 728-741.

Elam, J. J. and M. Mead (1990) "Can Software Influence Creativity?" *Information Systems Research* (1) 1, pp. 1-22.

Faure, C. (2004), “Beyond Brainstorming: The Effects of Different Group Procedures on Selection of Ideas and Satisfaction with the Process,” *Journal of Creative Behavior*, (38) 1, pp. 13-34.

Fern, E. F. (1982) "The Use of Focus Groups for Idea Generation: The Effects of Group Size, Acquaintanceship, and Moderator on Response Quantity and Quality," *Journal of Marketing Research* (19) 1, pp. 1-13.

Firestein, R. L. (1990), “Effects of Creative Problem Solving Training on Communication Behaviors in Small Groups,” *Small Group Research*, (21) 4, pp. 507-521.

Gallupe, B. R., L. M. Bastianutti, and W. H. Cooper (1991), "Unblocking Brainstorms", *Journal of Applied Psychology,* (76) 1, pp. 137-142.

Gallupe, R. B., W. H. Cooper, M. L. Grise, and L. M. Bastianutti (1994), "Blocking Electronic Brainstorms," *Journal of Applied Psychology,* (79) 1, pp. 77-86.

Gallupe, R. B., A. R. Dennis, W. H. Cooper, J. S. Valacich et al. (1992) "Electronic Brainstorming and Group Size," *Academy of Management Journal* (35) 2, pp. 350-369.

Gallupe, R. B., G. DeSanctis, and G. W. Dickson (1988) "Computer-Based Support for Group Problem-Finding: An Experimental Investigation," *MIS Quarterly* (12) 2, pp. 277-296.

Gallupe, R. B. and J. D. McKeen (1990) "Enhancing Computer-Mediated Communication: An Experimental Investigation into the Use of a Group Decision Support System for Face-to-Face Versus Remote Meetings," *Information & Management* (18) 1, pp. 1-13.

Garfield, M. J., J. T. Nolan, A. R. Dennis, and J. W. Satzinger (2001) "Modifying Paradigms-Individual Differences, Creativity Techniques, and Exposure to Ideas in Group Idea Generation," *Information Systems Research* (12) 3, pp. 322-333.

Gaustello, S. J. (1995), “Facilitative Style, Individual Innovation, and Emergent Leadership in Problem Solving Groups,” *Journal of Creative Behavior* (29) 4, pp. 225-239.

Gettys, C. F. (1987), “An Evaluation of Human Act Generation Performance,” *Organizational Behavior and Human Decision Processes*, (39) 1, pp. 25-51.

Graham, W.K. (1977), "Acceptance of ideas generated through individual and group brainstorming," *Journal of Social Psychology*, 101, 231-234.

Gryskiewicz, S. S. (1980) A Study of Creative Problem Solving Techniques in Group Settings. Doctoral Dissertation, University of London.

Hand, D.J. (1996) "Statistics and the theory of measurement. Journal of the Royal Statistical Society. Series A (Statistics in Society). (159) 3, 445-492.

Hender, J. M., D. L. Dean, T. L. Rodgers, and J. F. Nunamaker, Jr., (2002) "An Examination of the Impact of Stimuli Type and GSS Structure on Creativity: Brainstorming Versus Nonbrainstorming Techniques in a GSS Environment," *Journal of Management Information Systems* (18) 4, pp. 59-85.

Hu, L. T. and P. M. Bentler (1995) Evaluating Model Fit, in R. H. Hoyle (Ed.) *Structural Equation Modeling: Concepts, Issues, and Applications*, Thousand Oaks, CA: Sage, pp. 76-99.

Jackson, P. W. and S. Messick (1965) "The Person, the Product, and the Response: Conceptual Problems in the Assessment of Creativity," *Journal of Personality and Social Psychology* (33) 3, pp. 309-329.

Jessup, L. M., T. Connolly, and J. Galegher (1990), "The Effects of Anonymity on GDSS Group Process With Idea-Generating Task," *MIS Quarterly,* (14) 3, pp. 313-321.

Jessup, L. M., and D. A. Tansik (1991), "Decision Making in an Automated Environment: The Effects of Anonymity and Proximity with a Group Decision Support System," *Decision Sciences,* (22) 2, pp. 266-279.

Joreskog, K. G. and D. Sorbom (1988) *LISREL 7: A Guide to the Program and Applications*. Chicago, IL.: SPSS Inc.

Kelly, J. R. and Karau, S. J. (1993), "Entrainment of Creativity in Small Groups," *Small Group Research, (24) 2, pp. 179-198.*

Kirton, M. (1976), "Adaptors and innovators: a description of a measure," *Journal of Applied Psychology*, (61)5, pp.622-629

Kirton, M. (1987), "Adaptors and innovators: cognitive style and personality" in S.G. Isaksen (ed.), *Frontiers of Creativity Research, Buffalo, NY: Bearly Ltd., pp. 282-304.*

Kirton, M. (1989), *Adaptors and Innovators: Styles of Creativity and Problem Solving*, London: Routledge.

Kramer, M. W. and Kuo, C. L. (1997), "The Impact of Brainstorming Techniques on Subsequent Group Processes," *Small Group Research*, (28) 2, pp. 218-242.

Kramer, T. J., Fleming, G. P. and Mannis, S. M. (2001), "Improving Face-to-Face Brainstorming Through Modeling and Facilitation," *Small Group Research*, (32) 5, pp. 533-557. Not referenced in text

Lamm, H. and G. Trommsdorff (1973) "Group Versus Individual Performance on Tasks Requiring Ideational Proficiency (Brainstorming): A Review," *European Journal of Social Psychology* (3) 4, pp. 361-387.

Leggett Dugosh, K., Paulus, P. B., Roland, E. J. and Yang, H-C. (2000), "Cognitive Stimulation in Brainstorming," *Journal of Personality and Social Psychology*, (79) 5, pp. 722-735.

MacCrimmon, K. R. and C. Wagner (1994) "Stimulating Ideas through Creativity Software," *Management Science* (40) 11, pp. 1514-1532.

MacCrimmon, K. R. and C. Wagner (1991-92), "System for the Support of Alternative Generation," *Journal of Management Information Systems*, (8) 3, pp. 49-67.

Marakas, G. M. and Elam, J. J. (1997), "Creativity Enhancement in Problem Solving: Through Software or Process?" *Management Science*, (43) 8, pp. 1136-1146.

Massetti, B. (1996) "An Empirical Examination of the Creativity Support Systems on Idea Generation," *MIS Quarterly* (20) 1, pp. 83-97.

McGlynn, R. P., McGurk, D., Effland, V. S., Johll, N. L. and Harding, D. J. (2004), "Brainstorming and Task Performance in Groups Constrained by Evidence," *Organizational Behavior and Human Decision Processes*, (93) 1, pp. 75-87.

McLeod, P. L., Lobel, S. A. and Cox, T. H. Jr. (1996), "Ethnic Diversity and Creativity in Small Groups," *Small Group Research*, (27) 2, pp. 248-264.

Miura, A. and Hida, M. (2004), "Synergy Between Diversity and Similarity in Group Idea Generation," *Small Group Research*, (35) 5, pp. 540-564.

Mumford, M. D. (2001), "Tradeoffs Between Ideas and Structure: Individual Versus Group Performance in Creative Problem Solving," *Journal of Creative Behavior*, (35) 1, pp. 1-23.

Murthy, U. S. and D. S. Kerr (2003) "Decision Making Performance of Interacting Groups: An Experimental Investigation of the Effects of Task Type and Communication Mode," *Information & Management* (40) 5, pp. 351-361.

Nagasundaram, M. and R. P. Bostrom (1994-95) "The Structuring of Creative Processes Using GSS: A Framework for Research," *Journal of Management Information Systems* (11) 3, pp. 87-114.

Nunamaker, J. F., Jr., A. R. Dennis, J. S. Valacich, and D. R. Vogel (1991) "Information Technology for Negotiating Groups: Generating Options for Mutual Gain," *Decision Sciences* (37) 10, pp. 1325-1346.

Ocker, R., F. J., H. R., and K. Johnson (1998) "Effects of Four Modes of Group Communication on the Outcomes of Software Requirements Determination," *Journal of Management Information Systems* (15) 1, pp. 99-118.

Offner, A. K., Kramer, T. J. and Winter, J. P. (1996), "The Effects of Facilitation, Recording and Pauses on Group Brainstorming," *Small Group Research*, (27) 2, pp. 283-298.

Osborn, A. F. (1953) *Applied Imagination*, 1st Edition, New York: Scribner.

Osborn, A. F. (1963) *Applied Imagination*, 3rd Edition, New York: Scribner.

Parent, M., R. B. Gallupe, W. D. Salisbury, and J. M. Handelman (2000) "Knowledge Creation in Focus Groups: Can Technologies Help?" *Innovation and Management* (38) 1, pp. 47-58.

Paulus, P. B. and Dzindolet, M. T. (1993), "Social Influence Processes in Group Brainstorming," *Journal of Personality and Social Psychology*, (64) 4, pp. 575.

Paulus, B. P. and Yang, H-C. (2000), "Idea Generation in Groups: A Basis for Creativity in Organizations," *Organizational Behavior and Human Decision Processes*, (82)1, pp. 76-87.

Petrovic, O. and O. Krickl (1994) "Traditionally-Moderated Versus Computer Supported Brainstorming: A Comparative Study," *Information & Management* (27) 4, pp. 233-243.

Pinsonneault, A., Barki, H., Gallupe, R. B. and Hoppen, N. (1999), "Research Note: The Illusion of Electronic Brainstorming Productivity: Theoretical and Empirical Issues," *Information Systems Research*, (10) 4, pp. 378-380.

Plucker, J. A., Beghetto, R. A. and Dow, G. T. (2004), "Why Isn't Creativity More Important to Educational Psychologists? Potentials, Pitfalls, and Future Directions in Creativity Research," *Educational Psychologist*, (39) 2, pp. 83-96.

Potter, R. E. and Balthazard, P. (2004), "The Role of Individual Memory and Attention Processes During Electronic Brainstorming," *MIS Quarterly*, (28) 4, pp. 621-643.

Redmond, M. R., Mumford, M. D. and Teach, R. (1993), "Putting Creativity to Work: Effects of Leader Behavior on Subordinate Creativity," *Organizational Behavior and Human Decision Processes*, (55)1, pp. 120-151.

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

Rickards, T. (1992) Search Widely, Choose Wisely: A Case Study Exploring Structural and Behavioral Factors in Idea Generation and Evaluation Following Brainstorming on a Strategic Problem, Working Paper, Manchester Business School.

Roy, M. C., Gauvin, S. and Limayem, M. (1996), "Electronic Group Brainstorming: The Role of Feedback on Productivity," *Small Group Research*, (27) 2, pp. 215-247.

Runco, M. A. and R. E. Charles (1993) "Judgments of Originality and Appropriateness as Predictors of Creativity," *Personality and Individual Differences* (15) 5, pp. 537-546.

Santanen, E. L., R. O. Briggs, and G. J. de Vreede. (2000) The Cognitive Network Model of Creativity: a New Causal Model of Creativity and a New Brainstorming Technique. *Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000*.

Santanen, E. L., R. O. Briggs, and G. J. de Vreede (2004) "Causal Relationships in Creative Problem Solving: Comparing Facilitation Interventions for Ideation," *Journal of Management Information Systems* (20) 4, pp. 167-197.

Satzinger, J. W., M. J. Garfield, and M. Nagasundaram (1999) "The Creative Process: The Effects of Group Memory on Individual Idea Generation," *Journal of Management Information Systems* (15) 4, pp. 143-160.

Schweiger, D. M., W. R. Sandberg, and J. W. Ragan (1986) "A Comparison for Reducing Interval Overconfidence in Group Judgments," *Academy of Management Journal* (29) 1, pp. 51-71.

Shalley, C. E. (1991) "Effects of Productivity Goals, Creativity Goals, and Personal Discretion on Individual Creativity," *Journal of applied Psychology* (76) 2, pp. 179-185.

Shalley, C. E. (1995) "Effects of Coaction, Expected Evaluation, and Goal Setting on Creativity and Productivity," *Academy of Management Journal* (38) 2, pp. 483-503.

Shalley, C. E. and 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 and Human Decision Processes*, (84)1, pp. 1-22.

Shepherd, M. M., Briggs, R. O., Reinig, B. A., Yen, J. and Nunamaker, J. F. Jr. (1995-6), "Invoking Social Comparison to Improve Electronic Brainstorming: Beyond Anonymity," *Journal of Management Information Systems*, (12) 3, pp. 155-170.

Shirani, A. I. and M. H. A. Tafti (1999) "Task and Technology Fit: A Comparison of Two Technologies for Synchronous and Asynchronous Group Communication," *Information & Management* (36) 3, pp. 139-150.

Smith, G. F. (1998), "Idea-Generation Techniques: A Formulary of Active Ingredients," *Journal of Creative Behavior* (32) 2, pp. 107-133.

Sosik,J. J. , Avolio, B. J. and Kahai, S. S. (1997), "Effects of Leadership Style and Anonymity on Group Potency and Effectiveness in a Group Decision Support System Environment," *Journal of Applied Psychology*, (82) 1.

Sosik,J. J. , Avolio, B. J. and Kahai, S. S. (1998), "Inspiring Group Creativity: Comparing Anonymous and Identified Electronic Brainstorming," *Small Group Research*, (29) 1, pp. 3-31.

Straus, S. G. and J. E. McGrath (1994) "Does Medium Matter? The Interaction of Task Type and Technology on Group Performance and Member Reactions," *Journal of Applied Psychology* (79) 1, pp. 87-97.

Taylor, D. W., P. C. Berry, and C. H. Block (1958) "Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking?" *Administrative Science Quarterly* (3) 1, pp. 23-47.

Thornburg, T. H. (1991), "Group Size and Member Diversity Influence on Creative Performance," *Journal of Creative Behavior* (25) 4, pp. 324-333.

Torrance, E. P. (1965), *Rewarding creative behavior.* Englewood Cliffs, NJ: Prentice-Hall.

Treffinger, D.J. "The progress and peril of identifying creative talent among gifted and talented students," *Journal of Creative Behavior* (14) pp. 20-39.

Tung, L. L. and A. R. Heminger (1993) "The Effects of Dialectical Inquiry, Devil's Advocacy, and Consensus Inquiry Methods in a GSS Environment," *Information & Management* (25) 1, pp. 33-41.

Valacich, J. S., A. R. Dennis, and T. Connolly (1994) "Idea Generation in Computer-Based Groups: A New Ending to an Old Story," *Organizational Behavior and Human Decision Processes* (57) 3, pp. 448-467.

Valacich, J. S., A. R. Dennis, and J. F. Nunamaker, Jr. (1992) "Group Size and Anonymity Effects on Computer-Mediated Idea Generation," *Small Group Research* (23) 1, pp. 49-73.

Valacich, J. S., J.F. George, J. F. Nunamaker, Jr., and D. R. Vogel (1994) "Physical Proximity Effects on Computer-Mediated Group Idea Generation," *Small Group Research* (25) 1, pp. 83-104.

Valacich, J. S. and Shearer, T. (1995), "Devil's Advocacy and Dialectical Inquiry Effects on Face-to-Face and Computer-Mediated Group Decision Making," *Organizational Behavior and Human Decision Processes* (63) 2, pp. 158-173.

Valacich, J. S., R. Wachter, B. E. Mennecke, and B. C. Wheeler. (1993) Computer-Mediated Idea Generation: The Effects of Group Size and Group Heterogeneity. *Proceedings of the 26th International Conference on System Sciences, 1993*, pp. 152-160 IV.

Valacich, J. S., Wheeler, B. C., Mennecke, B. E. and Wachter, R. (1995), "The Effects of Numerical and Logical Group Size on Computer-Mediated Idea Generation," *Organizational Behavior and Human Decision Processes*, (62) 3, pp. 318-329.

Wagner, C. (1996) "Creative Behavior through Basic Inferences: Evidence from Person-Computer Interactions," *Journal of Creative Behavior* (30) 2, pp. 105-125.

Wong, Z. and Aiken, M. (2003), "Automated Facilitation of Electronic Meetings," *Information & Management*, (41) 2, pp.125-134.

Woodman, R. W., Sawyer, J. E. and Griffin, R. W. (1993), "Toward a Theory of Organizational Creativity," *Academy of Management Review*, Vol. 18, No. 2, pp. 293-321.

APPENDIX A. QUALITY SCALES

Table 21. Originality: The degree to which the idea is not only rare but is also ingenious, imaginative, or surprising.

Score	Level Description	Restaurant Examples	Tourism Examples
4	Not expressed before (rare, unusual) **And** Ingenious, imaginative or surprising; may be humorous	Buy out other surrounding restaurants. Have someone feed you the food while relaxing in a lawn chair by the pool. Play music that psychologically makes people hungry or thirsty.	Say that we have a religious relic like the Holy Grail
3	Unusual, interesting; shows some imagination	Have a roller derby night	Offer a special weekend visit to Rocky Point or Grand Canyon if you spend a week in Tucson
		Have individuals on campus passing out flyers and telling people about it, maybe have him/her wear something flashy	
2	Interesting	Use more spices, herbs and fresh ingredients to improve taste	Have more golf courses in town
		Entertainment that ranges from jazz to blues	
1	Common, mundane, boring	All-u-can eat salad bar for a nominal fee with the purchase of an entrée	Advertise

Table 22. Acceptability: The degree to which the idea is socially, legally, or politically acceptable.			
Score	**Level Description**	**Restaurant Example**	**Tourism Example**
4	Common strategies that violate no norms or sensibilities	Hand out flyers on campus	Advertise
		Offer healthy menu	
3	Somewhat uncommon or unusual strategies that don't offend sensibilities	Offer cool stories or jokes on the menu so it can be read while waiting	
		Telephones at each table, so you can talk from table to table	Give free watches to everyone
2	Offends sensibilities somewhat but is not totally unacceptable	Have crazy events through the night such as times when the bar tenders stand on the bar with a bottle of booze and walk down pouring it into different mouths	Make a brochure with the "Wild Men"
		Allow patrons to dance on the tables	
1	Radically violates laws or sensibilities or Totally unacceptable business practice.	Put some addictive substance in the food and milk the students for everything they have	Burn the town down and start again
		Use the same grease for the next month to cook fries, chicken nuggets, and other health foods	

Table 23. Implementability: The degree to which the idea can be easily implemented.			
Score	**Level Description**	**Restaurant Example**	**Tourism Example**
4	Easy to implement at low cost or non-radical changes	Have different varieties of music on certain nights	Have a list of things to do under $15
		Sometimes have people selling your food on the mall or sponsoring stuff around campus	
3	Some changes or reasonably feasible promotions or events	Have a grand re-opening with a radio station, with free food, prizes and contests. Make sure there is lots of advertising in and around the university in conjunction with the community	Have a lot more advertising
		Make the restaurant honour all-aboard cards and make it so the students receive an extra 10% off food purchases if they use all aboard	
2	Significant change or expensive or difficult but not totally impossible to implement	Remodel the restaurant in an up to date style	Build a great art museum
		Pay beautiful people to eat there so others will want to as well	
1	Totally infeasible to implement or extremely financially nonviable	Free lunch on every Friday of the week	Create an ocean
		Convince the professors to give the students extra credit for going to the restaurant	

Table 24. Applicability: The degree to which the idea clearly applies to the stated problem.			
Score	**Level Description**	**Restaurant Example**	**Tourism Example**
4	Solves an identified problem that is directly related to the stated problem (do X to get Y, and Y is part of the stated problem)	Hire both English and Spanish speaking employees for a broader base for customers	Research to find target market for tourists
		Work with restaurants around you in order to jointly draw more customers to your area	
3	Solves an implied problem that is related to the stated problem (do X to get an implied Y, which applies to the stated problem)	Free lunch on every Friday of the week	Have a lot more advertising
		Increase variety of the drinks menu	
2	May have some benefit within a special situation and somehow relates to the stated problem (do X, which somehow relates to the stated problem)	Have an attendant in the bathroom to help with cologne and mouthwash	Build indoor skiing facilities
		Have the Christmas coloured mints from December to January	
1	Intervention is not stated or does not produce a useful outcome (no X) or (do X for useless Y)	Put the restaurant in a bad location and car theft will free up parking space	Tell tourists to bring bottled water
		Lobby congress for lower taxes to provide cheaper food	

Table 25. Effectiveness: The degree to which the idea will solve the problem.			
Score	**Level Description**	**Restaurant Example**	**Tourism Example**
4	Reasonable and will solve the stated problem without regard for workability (If you could do it, it would solve the main problem)	Buy out other surrounding restaurants so people will stay at your place	Say that we have a religious relic like the Holy Grail
		Put some addictive substance in the food and milk the students for everything they have	Research to find target market for tourists
3	Reasonable and will contribute to the solution of the problem (It helps, but it is only a partial solution)	Provide birthday specials. Perhaps a free meal for the birthday person.	Advertise the sunsets
		Use more spices, herbs and fresh ingredients to improve taste	
2	Unreasonable or unlikely to solve the problem (It probably will not work)	Have crazy events throughout the night such as times when the bar tenders stand on the bar with a bottle of booze and walk down pouring it into different mouths	Believe in Tucson - think of positive things
		Put a full court basketball facility in the back	
1	Solves an unrelated problem (It would not work, even if you could do it)	Have employees that can speak English	Tell tourists to bring bottled water
		Free fighting	Do not advertise - make them find their own entertainment
		Don’t put the sign up ‘Ketchup upon request’	

Table 26. Completeness: The number of independent subcomponents into which the idea can be decomposed, and the breadth of coverage with regard to who, what, where, when, why, and how.

Score	Level Description	Restaurant Example	Tourism Example
3	Comprehensive, with three or more parts from at least two of the 5 Ws + H (who, what, why, when, where, how), e.g. (what + when + where) or (what + what + why)	Advertise that slow is better - results in more care taken and fresh food used	Have observatory on campus open July through September for free viewing at night
		Owner should ask people on campus what they've heard about the restaurant and improve on criticism	The city should get together with a hotel chain and an airline chain to put together a package deal for tourists
2	Contains two parts from different dimensions (5 Ws + H), such as, but not limited to (what + where), (what + why), (what + how), *or* three or more parts of only one of the 5 Ws + H (e.g., what + what + what)	Hand out flyers on campus	Have camping tour company - all you bring is yourself and they provide everything else
		Free lunch on every Friday of the week	
1	Contains one or two parts from the same dimension and usually the "what" (e.g., (what) or (what + what))	Create a breakfast menu	Advertise
		Provide free parking	

Table 27. Implicational explicitness: The degree to which there is a clear relationship between the recommended action and the expected outcome.

Score	Level Description	Restaurant Example	Tourism Example
3	Implication is clearly stated and makes sense (do X so that Y)	Fix up the place to attract more people, people don't like to go someplace that looks bad	Have creative attractions to make it easier to visit
		Decorate the place colourfully so it stands out from the rest so it catches the viewers' eyes as they drive by	
2	Implication is not generally accepted or is vaguely stated (do X, which solves a not-generally-accepted Y) *or* (do X which solves a vaguely stated Y)	Have a frequent meal plan where the more you come in the more free food you get	Promote health and fitness - build a health spa
		Advertise in an inventive way that will bring in better people	
1	Implication is not stated, even though relevant (do X without a stated Y)	Entertainment that ranges from jazz to blues	Advertise
		Add a buffet	

Table 28. Clarity: The degree to which the idea is clearly communicated with regard to grammar and word usage.			
Score	**Level Description**	**Restaurant Example**	**Tourism Example**
3	Crisp, with standard usage, including complete sentences or well-developed phrases, and every word is commonly understood	Stay open late during finals and offer cheap coffee	Advertise in Canada to senior citizens
		Create a breakfast menu	
2	Understandable, with acceptable usage or understandable phrases; some words might be known only within a small context; sentences might contain fragments or be incomplete (yet understandable)	Offer different food than at the union	Have mass transit for old people
		Offer a few of the same thing - only pizza, or sandwiches, etc.	Low level of prejudice
1	Vague or ambiguous words or use of poor language structure	Advertise in an inventive way that will bring in better people	Build up the freeway system here
		Have good deals on your menu	
		Value packs	

APPENDIX B. PARADIGM RELATEDNESS

Table 29. Paradigm relatedness: The degree to which an idea preserves or modifies a paradigm. PM ideas are sometimes radical or transformational.

Score	Restaurant Problem	Tourism Problem
4	Paradigm Breaking. Introduces new elements and changes the relationship with the customer. Also includes any ideas that focus on the larger problem for the restaurant—staying in business (e.g., different way to make money, selling the restaurant, etc.). Examples: Spread nasty rumors about the other restaurants in the area. Put roaches in other restaurants' kitchens and make sure customers find them.	Paradigm Breaking. Introduces new elements and changes the relationship with the tourists. Includes more radical reasons for visiting Tucson, such as anything that would make Tucson famous. Example: Advertise that California will fall into the ocean—"Get beach-front properties."
3	Paradigm Stretching. Changes the relationship with the customers (i.e., gives them something other than food to attract them to the restaurant.) Also includes research. Examples: Have a roller derby night. Put a full-court basketball facility in the back.	Paradigm Stretching. Changes the relationship with the tourists. Includes other reasons for visiting Tucson, such as health, education, business, etc., and changes to Tucson itself. Also includes research. Example: Advertise that Tucson helps health problems like arthritis.
2	Slightly Paradigm Stretching. Introduces new elements (e.g., different food, different hours, different decor, different ways of advertising, etc.) but still focuses on serving food. Examples: Use more spices, herbs, and fresh ingredients to improve taste. Stay open late during finals and offer cheap coffee.	Slightly Paradigm Stretching. Introduces new elements (e.g., different ways of advertising, new attractions, etc.), but still focuses on vacations. Example: Use Internet for advertising.
1	Paradigm Preserving. Serving food to students. Example: Hand out flyers on campus.	Paradigm Preserving. Usual ways of attracting tourists to Tucson for vacations. Example: Advertise.

APPENDIX C. TRAINING INSTRUCTIONS

This appendix is intended to help train raters in the use of the scales presented in this research. The information provided in this paper provides a serviceable foundation for training raters. We recommend that training be conducted in the following sequence.

Construct Definitions. This paper defines each construct and subconstruct. Raters should read the first part of Section II where we define, and distinguish between, idea quality, idea novelty, and idea creativity and where we show how creativity has been measured in the past. Next, raters should read the subsection titled CLARIFICATION OF DIMENSIONS AND SUB-DIMENSIONS in Section III, including Table 8, where each dimension and sub-dimension is explained and clarified.

Counting Protocol. Raters should also understand and apply the idea-counting protocol defined in Appendix D

Level Descriptions. Raters should also read and understand the textual descriptions of each ordinal level in each scale, as shown in Appendices A and B. These appendices not only define each ordinal level but also provide multiple examples of ideas from both of our data sets that corresponded to each ordinal level. Researchers will have to define a paradigm-relatedness scale for the idea generation task that they use because that which qualifies as a common element of the solution and the relationship between the elements are problem-specific.

Inter-Rater Reliability. Lastly, for each scale, raters need to train on randomly selected sets of 50 to 100 ideas to hone consistency. This allows raters to examine the reasons for their consistent and inconsistent ratings. This training refinement process should be done on successive samples of ideas until reasonable inter rater reliability (Cronbach's alpha approx. >0.70) is achieved.

APPENDIX D. IDEA COUNTING APPROACHES

The task of counting ideas is essential in ideation studies. Many researchers, however, do not provide explicit details concerning their specific idea counting approach. Of the 90 articles included in our sample, only 20 (22%) explained their idea counting method, a significantly small minority (Chi-square = 27.78, p. = 0.000). Moreover, some descriptions included only a sentence or two while a few provided more detail.

For example, many researchers indicate that their pool of ideas was coded independently by two or more raters, who identified and counted the number of unique ideas. A certain level of agreement or intercorrelation among the raters concerning the coding of the data is then usually taken as an indicator of adequate inter-rater reliability (Campbell, 1968; Dennis, 1994; Easton, Easton, & Belch, 2003; Paulus & Yang, 2000; Shirani, Tafti, & Affisco, 1999; Valacich, Dennis, & Connolly, 1994). But to accomplish a high inter-rater reliability, an implicit or explicit counting approach is required.

Variable or undisciplined counting methods may lead to inconsistencies either within or across studies. This can lead to difficulties involving the interpretation and comparability of findings across studies. But counting ideas is not necessarily trivial. For example,

consider the following ideas generated in responses to the question of how to improve business at a local restaurant)[8]:

1. Do something to attract more people to the restaurant.
2. Run a happy hour.
3. Advertise more.
4. Advertise in the newspaper.
5. Offer a reduced price menu at a certain time.
6. Offer a reduced price menu.
7. Have different prices for different times of the day.

The list above illustrates several potential difficulties involved with counting ideas. For example, idea #2 and idea #5 are very similar to one another. Idea #4 is an elaboration of ideas #3. In turn, ideas #6 and #7, when taken together, are very similar to individual ideas #2 and #5. Without a specific and clear procedure, it becomes difficult to determine the quantity of unique ideas that are present in any given set of ideas.

In recognition of the problems inherent in assessing the quantity of ideas that exist in any given pool of responses, researchers have proposed more formal methods that include approaches for dealing with duplicates, specializations of other ideas, highly ambiguous, and irrelevant ideas. These are dealt with at the individual or group level since these are common units of analysis in ideation research.

An elaborate protocol for determining the quantity of ideas present in any given pool was developed by Bouchard and Hare (1970). This approach has been adopted in studies conducted by Gallupe, et al., 1991; Gallupe, et al., 1992; Garfield, et al., 2001; Barki and Pinsonneault, 2001; and Valacich, et al., 1992. It is now combined with steps from Connolly et al. 1993) to provide an overall approach.

1. Remove statements that do not specifically address the problem task (Bouchard and Hare, 1970).
2. Remove ideas that are general or ambiguous to the extent that their specific intent or impact cannot be determined (Bouchard and Hare, 1970).
3. Remove duplicates (Connolly, et al. 1993).
4. If the same individual or group produces both general and specific versions of an idea, only the specific version is counted. Thus in the restaurant example above, idea #4 would count but idea #3 would not (Connolly, et al. 1993).
5. If a general rule or statement is followed by a list of examples, the general rule together with the first example is counted as an idea. Each subsequent example from the list that is also explained in detail is also counted as a unique idea (Bouchard and Hare, 1970).
6. Remove ideas that resulted from misunderstandings of the problem or task if they do not have a direct analog in the actual task. For example, using the well-known Thumbs task, a subject may misinterpret the task such that they believed each person now had 11 fingers rather than 12 (as intended). Thus, switching to a base 11 numeric system would count as a unique idea since its

[8] The authors thank an anonymous reviewer for providing this excellent example demonstrating the potential difficulties of counting ideas.

> analog in the actual task is switching to a base 12 numeric system. If no such analog exists, the idea is not counted (Bouchard and Hare, 1970).

For researchers who would like to use the infrequency method for identifying unique or rare ideas, it is necessary to determine the frequency of ideas across all groups or individuals. This approach can be done after formal method of identifying redundancies (such as those indicated above) has been performed. Gettys et al. (1987) provides a convenient way of organizing ideas into a tree structure that can aid in this analysis. This approach characterizes "performance by structuring the actions generated by the subject and comparing the subject's structure with a more complete 'treelike' act structure created by the experimenter. The limbs of the tree consist of the generic types of actions that could be taken to solve the problem, its branches are major variations of these generic actions, and its twigs minor variations of the branches" (Gettys, et al., 1987, p27). Using this mechanism, an individual subject would not be expected to duplicate the entire tree structure. However, excellent performance would be indicated when the individual's (or group's) tree structure included certain important limbs and high utility branches. The "optimal" tree structure could be developed by the experimenter or could be the result of pooling responses from a sufficiently large set of subjects until adding new subjects does not add anything significant to its structure. The actual performance measure then derives from examining the extent to which subjects generate major structural elements of the optimal tree.

In summary, research would be more comparable if an explicit, consistent counting approach is used. In our sample, the twenty articles that explain their idea counting methods describe a total of seven different counting mechanisms. The counting methods employed by the remaining 70 studies were not reported. We have drawn from two of the more detailed methods (Connolly et al. 1993 and Bouchard and Hare, 1970) to provide a useful set of conventions. Future research could compare different counting conventions.

About the Authors

Douglas L. Dean is an Associate Professor at the Marriott School of Management at Brigham Young University. He received his Ph.D. in MIS from the University of Arizona in 1995. Dr. Dean's research interests include electronic commerce technology and strategy, requirements analysis, and collaborative tools and methods. His work has been published in Management Science, Journal of Management Information Systems, Data Base, Communications of the AIS, Expert Systems with Applications, Group Decision and Negotiation, and IEEE Transactions on Systems, Man, and Cybernetics.

Jillian M. Hender is a member of the Associate Faculty at Henley Management College, UK. She received a B.Sc. in Physics from University College, London, and a Ph.D. and MBA from Henley Management College. She has worked in the computer industry as an analyst/programmer and consultant. Her research interests include creativity, innovation management, teams, group support systems, and information technology for collaborative work. Her work has been published in the Journal of Management Information Systems and the Journal of General Management.

Thomas L. Rodgers is an independent consultant. He received his Ph.D. in MIS form the University of Arizona in 1999. Dr. Rodgers' research interests include information technology applications within healthcare contexts; team aspects of software engineering; e-service technology and diffusion; and learning, expertise and creativity in collaborative environments.

Eric L. Santanen is an Assistant Professor of Management at Bucknell University. He earned his Baccalaureate and Mater's degrees in Computer Science at the New Jersey Institute of Technology and his Doctorate in Management Information Systems at the University of Arizona. His primary research interest deals with developing techniques and technology that can help individuals and groups generate creative solutions to complex problems. This work is based upon foundations of human cognition, problem solving, group support systems, and creativity and has been presented before numerous international audiences. Other research interests include human computer interfaces and technology-supported learning.

Journal of the Association for Information Systems
ISSN: 1536-9323