Master's Theses   Theses and Dissertations

2014

The Effect of Semantic Clustering on Idea Quality in Individual and Group Ideation

Amanda Christine Egan
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

Part of the Social Psychology Commons

LOYOLA UNIVERSITY CHICAGO

THE EFFECT OF SEMANTIC

CLUSTERING ON IDEA QUALITY IN

INDIVIDUAL AND GROUP IDEATION

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF ARTS

PROGRAM IN APPLIED SOCIAL PSYCHOLOGY

BY

AMANDA C. EGAN

CHICAGO, ILLINOIS

DECEMBER 2014

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Rietzschel, Nijstad, and Stroebe (2007) have demonstrated the benefits of "deep exploration" on creative idea generation. The current study attempted to refine this understanding by differentiating whether this effect is due simply to the number of ideas generated within a specific semantic category (fluency) or the way in which semantic categories are explored (clustering). Four conditions compared maximum versus minimum clustering crossed with nominal and interacting groups, with total quantity and fluency held constant. The unique effects of these manipulations on the total number of high-quality ideas generated, as well as the specific number of highly-original and highly-feasible ideas generated during brainwriting sessions were measured. The results provide tentative support for the idea that over the course of a brainstorming session, a minimum clustering paradigm is better suited toward generating more high-quality ideas and more highly-original ideas.

CHAPTER ONE

INTRODUCTION

A large body of research spanning almost sixty years has evaluated the effectiveness of an idea generation intervention known as brainstorming (Larson, 2010). Brainstorming is an intervention to assist groups in creative idea generation. It was first introduced by an advertising executive, Alex Osborn, in his 1953 book *Applied Imagination*. He suggested that four key principles be followed: a) the primary focus should be on quantity rather than quality, b) free-wheeling and the suggestion of unusual ideas should be encouraged, c) members should attempt to combine and build on one another's ideas, and 3) there should be no criticism of ideas. Osborn believed that by following these rules, a group would be able to generate twice as many ideas as its individual members would have had they worked alone (Nijstad & Stroebe, 2006).

The majority of research surrounding brainstorming has focused on a specific dependent variable – global productivity. Global productivity refers to the total number of ideas generated during a brainstorm session. This focus has typically been justified by the widely held assumption that "quantity yields quality." Because the aim of practitioners in using brainstorming is typically the creation of high-quality ideas (rather than just a high volume of ideas), manipulations aimed at increasing the quantity of ideas are only worthwhile if they in fact do also result in more high quality ideas. But, quantity

1

may be an unnecessarily cumbersome route to generating a few high-quality ideas. As such, researchers are calling for the development of theories and methods aimed at improving idea quality more directly rather than simply by inflating quantity (Reinig & Briggs, 2008). One study that heeded this call was conducted by Rietzschel, Nijstad, and Stroebe (2007). They found that deep exploration within a semantic category had beneficial effects on the quality and creativity of ideas produced within that category. The current study will attempt to better understand the underlying cognitive mechanisms at work during "deep exploration" in order to better refine this tool for improving idea quality.

Defining and Measuring Quality

A broad but generally accepted definition of quality states that it is, "some combination of *originality* (the degree to which an idea is innovative) and *appropriateness* (e.g., the degree to which a product or an idea is relevant to the topic, or is thought to be practically feasible.)" (italics in original) (Rietzschel, et al., 2007) Thus, idea quality reflects a combination of desirable factors such as originality and feasibility.

There are multiple ways to assess the quality of a set of ideas, including idea-count, sum-of-quality, average quality, and good-idea-count measures (Reinig, et al., 2007). A simple idea-count uses quantity as a surrogate for measuring quality when the correlation between the two is high (Diehl & Stroebe, 1987). However, it would be preferable to measure quality directly. Sum-of-quality scores are obtained by assigning a quality rating[1] to each idea and then totaling all of the quality scores. The main problem

---

[1] Often the ratings are done on either a five-point or seven-point scale where higher numbers indicate higher quality.

with sum-of-quality scores is that they are biased in favor of greater quantity. More ideas yield a higher sum-of-quality score even if none of the ideas are actually high quality. Average quality is also obtained by assigning a quality rating to each idea, but here the ratings are then averaged. Unfortunately, averaging is not as useful as it might initially seem. For instance, consider two hypothetical brainstorming sessions in which ideas are rated using a five-point scale, with higher numbers indicating higher quality. If Session 1 yields ideas with the scores 1, 1, 1, 5, 5, and 5, and Session 2 yields ideas with the scores 3, 3, 3, 3, 3, and 3, then both session would result in an average quality score of 3. Thus, average quality is not useful from the perspective of a practitioner trying to decide which session generated more high quality ideas. Finally, a good-idea-count also uses ratings based on a scale. In this case however, only those ideas meeting a minimum quality criterion are counted (e.g., ideas with a quality rating of four or five on a five-point scale) (Reinig, et al., 2007). This measure does not penalize groups for the inclusion of poor-quality ideas, as the average quality measure does, and does not advantage groups that include many low quality ideas, as the sum-of-quality score does. Good-idea-counts would therefore appear to be the best measure of quality, and so are emphasized in this study. Thus, wherever the dependent variable "quality" is mentioned here, it can be assumed that it is in reference to a good-idea-count, unless otherwise specified. Furthermore, attention will be paid mainly to those studies that have used a good-idea-count to measure the dependent variable, quality.

The Quantity-Quality Relationship

While quantity and quality[2] do seem to be strongly positively correlated ($r = .82$, Diehl &

Stroebe, 1987), work by Rietzschel et al. (2007) suggests that it is quantity within a

particular semantic category that is most beneficial to generating more high-quality ideas.

Manipulations used in other studies have also helped brainstormers generate more high

quality ideas, but these have operated by increasing global productivity without being

able to stimulate the generation of more high-quality ideas directly.

Diehl and Stroebe (1987) conducted an experiment in which they manipulated

type of session (individual vs. group) and type of assessment (personal vs. collective),

which resulted in four conditions. Both individuals and four-person interacting groups

recorded ideas verbally into lapel microphones. Those in the personal assessment

condition were told that their performance would be compared with another individual's

performance whereas, those in the collective assessment condition were told that their

group's performance would be compared with another group's performance. Ideas were

rated on two five-point subscales – originality and feasibility. On their scales, lower

numbers indicated better scores. A "good idea" was defined as one that was scored a 1

on one subscale, and no more than a 2 on the other subscale. Diehl and Stroebe (1987)

found that the total number of ideas was highly correlated with the number of good ideas

($r = .82$), but the average quality of the ideas was not affected by their manipulations.

This led them to conclude that their manipulations increased both good and poor quality

ideas. This illustrates the potentially cumbersome nature of using quantity to yield

---

[2] Measured as a good-idea-count determined using a similar cut-off method described above and used in the current study.

quality, and the need for additional research to investigate methods that more directly yield high-quality ideas. From a practitioner's point of view, an intervention that results in more high-quality ideas without also increasing poor-quality ideas would be preferable.

A study that used a manipulation with this aim was conducted by Parnes and Meadow (1959). They asked some participants to generate only good ideas, while asking other participants to generate as many ideas as possible. The authors found that specifically asking participants to produce "good" ideas (referred to as "non-brainstorming instructions" in their study) yielded fewer good ideas than instructions emphasizing the quantity of ideas (referred to as "brainstorming instructions" in their study).

However, the nature of the Parnes and Meadow (1959) experimental design leaves unclear the cause of their finding. Ideas were rated on two subscales – uniqueness and value. Each idea was scored on a three-point scale, where higher numbers indicate more uniqueness or value. In order for an idea to be considered "good" it had to receive a combined score of five, meaning it could not receive a score of 1 on either subscale, and had to receive a score of 3 on at least one of the scales. However, the "good idea" instructions read: "You are to list all the *good* ideas you can think up. Your score will be the total number of good ideas. Don't put down any idea unless you feel it is a *good* one" (italics in original; p. 173). The problem with this is that the term quality, and possibly the term "good" as well, may be too vague to prompt the generation of novel ideas (Runco, Illies, and Reiter-Palmon, 2005). Therefore, specifically prohibiting participants

from recording an idea unless they believed it to be "good," without providing

clarification that "good" would be determined by uniqueness and valuableness, may have

prevented participants from recording unique ideas. In other words, the instructions may

have inadvertently encouraged conventional ideas, thereby artificially diminishing the

number of good ideas generated using these instructions.

Another difficulty created by the "good-idea instructions" is that participants in

the Parnes and Meadow (1959) study were actually being asked to perform two distinct

cognitive tasks. First, they had to generate ideas. Second, they had to evaluate the

"goodness" of those ideas and decide whether or not to record them. This is twice the

amount of cognitive work asked of those using the quantity instructions, who were

specifically told, "Forget about the quality of ideas entirely. Express any idea which

comes to your mind." (p. 173).

Accordingly, the claim that Parnes and Meadow (1959) make that "more 'good'

ideas were produced under the brainstorming instructions than under the non-

brainstorming instructions" (p. 175) may be due to an imbalance between conditions

created by the wording and goals implied by the instructions. In fact, the correlations

between the number of "good" ideas and global productivity obtained in the non-

brainstorming/quality instruction conditions ($r = .64$ and $.81$) were not very different

from those in the brainstorming/quantity instruction conditions ($r = .67$ and $.71$). A better

manipulation, then, would be to somehow encourage high-quality ideas without asking

participants to perform two cognitive tasks. This may be how deep exploration works.

By encouraging quantity within a semantic category, quantity is still the only cognitive goal and quality is the natural byproduct of the process.

Paulus, Kohn, and Arditti (2011) point out one final flaw in Parnes and Meadow's (1959) study. Because there was no control group without specific quantity or quality instructions, it is unclear whether the quantity instructions improved performance or the quality instructions inhibited performance. Thus, Paulus et al. (2011) designed a study to compare four types of instructions: instructions with a quantity emphasis, instructions with a quality emphasis, instructions with a combined quality and quantity emphasis, and a baseline control condition in which instructions had neither a quantity nor a quality emphasis. They found that those in the quantity-focused instruction condition outperformed the other three conditions in both total number of ideas and number of high-quality ideas. High-quality ideas were defined in terms of novelty (uniqueness) and utility (positive impact) using two five-point scales[3]. They found that the quality instruction condition did not differ significantly from the control condition. This led them to conclude that it was a benefit due to the quantity instructions, rather than diminished performance in the quality condition, that was responsible for the differences between these two conditions. These findings reinforce the idea that a quantity-emphasis is the key to generating many high-quality ideas, and that perhaps encouraging quantity within a specific semantic subcategory may be a particularly promising intervention.

In an attempt to better understand the quality-quantity relationship, Reinig and Briggs (2008) observed fourteen interacting brainstorm groups. They demonstrated empirically a curvilinear relationship between quantity and quality, with a positive but

---

[3] Specific cut-off scores to qualify as "high-quality" were not provided in the article.

decreasing slope (rather than a simple linear slope). These authors postulated that the observed diminishing returns of increased quantity for producing greater quality was the result of two separate phenomenon, cognitive inertia and solution space limitations. Cognitive inertia occurs when spreading activation of problem-relevant information eventually makes less relevant information accessible and there is a lack of external stimuli to aid in switching to a more useful line of thought. Solution space limitation refers to constraints imposed by the number of viable solutions for a given problem. Thus, Reinig and Briggs (2008) suggest that the relation between quantity and quality may depend on factors related both to the problem or task itself as well as the person or people generating the solutions. Perhaps then, a better intervention can be designed that takes into account these factors.

Clustering and the Search for Ideas in Associative Memory

As previously mentioned, a manipulation that is less dependent on global productivity and has been shown to have an impact on idea quality is deep exploration within semantic categories (Rietzschel et al., 2007). To understand deep exploration, the concepts of fluency and clustering must first be discussed, and these are best understood in the context of the Search for Ideas in Associative Memory (SIAM) model of idea generation. Simplified, SIAM is a theory of idea generation built on Raaijmakers and Shiffrin's (1981) Search of Associative Memory (SAM) model. SIAM assumes the cognitive mechanisms of long-term and working memory systems. It suggests that idea generation is a two-loop process. First, an image retrieval loop occurs where problem-relevant cues are used to retrieve an image (a concept and its related information) from

long-term memory. Once an image is successfully retrieved, an idea production loop

occurs in which the retrieved image is used in working memory to generate new ideas.

The idea production loop is executed over and over again until no new ideas can be

generated. When this happens, a phenomenon known as cognitive failure, the image

retrieval loop is reactivated and the process repeats. Image retrieval is thought to be a

relatively time consuming, effortful process, whereas, idea generation from an image is

thought to be relatively fast and somewhat automatic (Nijstad & Stroebe, 2006).

As noted above, an image may be thought of as the cue and its related information

stored together in LTM. It is assumed that the information contained in the image is

mostly semantically related (Nijstad & Stroebe, 2006). Thus, when individuals generate

ideas using an image, those ideas should tend to be produced in a series of semantically

related ideas known as "clusters" (Rietzschel et al., 2007). The term "clustering" thus

refers to the natural tendency of semantically related ideas to be grouped together in time,

separated by shorter intervals than those found between semantically dissimilar ideas.

The result is an idea stream that can be pictured as follows: $A_1A_2A_3 \rightarrow B_1B_2 \rightarrow$

$C_1C_2C_3C_4 \rightarrow B_3B_4B_5$. In this idea stream, three semantic categories are sampled,

represented by the letters. The subscripts identify unique ideas within a semantic

category. Fluency refers to the total number of ideas per category, so B is the category

with the highest fluency (five ideas in this example). Clustering, on the other hand, refers

to the number of consecutive ideas from a given semantic category, so the third cluster is

the longest cluster (four ideas in this example). Because the idea generation loop is more

automatic than the image retrieval loop, successive ideas can be generated faster from the

same category than from different categories. Thus, SIAM predicts that more clustering is more efficient and more productive in terms of generating a high quantity of ideas in a fixed length of time. However, efficiency relates to global productivity and does not speak directly to the impact of clustering on idea quality in this model (Nijstad & Stroebe, 2006). But, work by Rietzschel et al. (2007) moves in the direction of helping to make such a prediction.

Rietzschel et al. (2007) demonstrated that deep exploration of a category is required in order to get past the highly accessible ideas and to work toward ideas that otherwise might not have surfaced. Because quality is a function of not only feasibility but originality, the less accessible ideas are likely to be more original and therefore have the potential to be higher quality ideas. To demonstrate this effect, Rietzschel et al. (2007) conducted an experiment using 93 University of Amsterdam psychology students. The researchers primed a relevant solution subcategory (i.e., nutrition, hygiene, or sports) for the problem "What can people do to improve or maintain their health?" They then had participants generate ideas. They found that participants generated more ideas within the primed subcategory (compared to the number generated in the un-primed categories), and that those ideas had a higher average originality and contained a higher percentage of high-quality ideas (defined as highly original and highly feasible), than did the ideas in other semantic categories. So, if a participant had received the prime "nutrition," he or she generated more nutrition-related ideas than sports- or hygiene-related ideas, and the nutrition-related ideas had a higher average originality and contained a higher percentage of high-quality ideas than did the ideas generated from the other categories. Because

they generated more ideas within the primed category, it is believed that they explored the primed category more deeply, and that this is what produced the higher quality and originality within that category. So, it was not quantity per se, but quantity within a semantic category that proved beneficial.

Continuous vs. Intermittent Depth of Exploration

There are at least two ways that depth within a semantic category can be attained: (a) working for a prolonged period of time within the category, and (b) revisiting the category frequently throughout an idea generation session. The same amount of fluency can be achieved either way, but the average cluster length would be higher using the first approach. While Rietzschel et al. (2007) demonstrated that deep exploration of a semantic category led to a greater percentage of high-quality ideas within that category, it is unclear whether their results are due to increased category fluency, to more clustering, or to both.

The Current Study

To parcel out the possible causes, the current study included two conditions wherein fluency was held constant but the amount of clustering was experimentally varied. Participants used a brainwriting procedure to generate ideas within three semantic categories. Brainwriting is a form of brainstorming where participants write their ideas rather than share them verbally. This is often done in interacting groups where index cards are used to record and share ideas between members. Using the brainwriting task, two levels of clustering were experimentally manipulated – maximum clustering and minimum clustering. In the maximum clustering condition, participants worked from one

semantic category until they had generated six ideas before being presented with the next semantic category prompt. In the minimum clustering condition, participants rotated between the three semantic categories in round-robin fashion until they generated six ideas in each one. Thus, the difference between these conditions is the average cluster length (number of ideas per cluster). The average cluster length in the maximum clustering condition is approximately six, whereas the average cluster length in the minimum clustering condition is approximately one. There appear to be both benefits and challenges associated with each condition that support two alternate hypotheses. I will discuss the support for each in turn.

Theoretically, if a given semantic category is explored continuously, a single image can be retrieved and used without interruption. This is because working within a semantic category should allow for the use of a single image, whereas switching between semantic categories likely requires retrieving different images. It has been argued that less switching between categories prevents cognitive interference (Nijstad, et al., 2002). Cognitive interference occurs when the ideas of another member disrupt one's own cognitive process of idea generation. The interference is thought to come not simply from exposure to another person's ideas, but from exposure to ideas that are drawn from a different category than one's own current ideas, thus disrupting one's own cognitive process (Nijstad et al., 2002). So in the maximum clustering condition of the current study, where members generated successive ideas from the same semantic category, there should be relatively little cognitive interference.

There is the possibility that more cognitive failure will be experienced by those in the maximum clustering condition because they will be prevented from switching to a new semantic category when they experience difficulty generating a new idea from the current one. Interestingly, this may lead to even more original ideas, and thus possibly higher quality ideas. This phenomenon has been labeled the *burden effect* (Burton, 1987). This effect has been observed in studies that used a multi-round procedure wherein participants worked individually at first and then shared and voted on ideas as a group. When participants experienced cognitive failure (were unable to contribute a new idea) they had to announce a "pass" during the idea sharing step. Researchers observed that participants seemed embarrassed at their inability to contribute. When it came time for that participant to contribute during the next round, the idea they shared was typically much more creative. Researchers believed that this is because the participants spent the interim in "intense mental activity" (Burton, 1987). If those in the maximum clustering condition in the current study are forced to push through cognitive failure rather than switch semantic categories, subsequent ideas may be even more original than before, possibly resulting in more high-quality ideas. Thus there is evidence for the following hypothesis:

- *H_1: More high-quality ideas will be generated by those in the maximum clustering condition than by those in the minimum clustering condition.*

Conversely, there are two conceivable challenges to working continuously from a single semantic category. First, when participants are required to generate successive ideas from a single semantic category, knowledge pertinent to other categories might

become less accessible, thus making it hard for them to switch to another category when it is time to do so (as they were required to do in this study) (Rietzschel, et al., 2007). Second, higher incidents of cognitive failure, the inability to generate an additional idea, should occur with maximum clustering because participants are prevented from switching when they naturally might be inclined to do so (Nijstad & Stroebe, 2006). Nijstad and Stroebe (2006) found more cognitive failure to correlate negatively with task enjoyment, satisfaction, and participants' expectation of being able to generate more ideas. They did not make a comparison with idea quality, but if satisfaction, enjoyment, and expectations of success are needed for persistence at a task, then perhaps greater cognitive failure will have a negative impact on the generation of high-quality ideas.

Therefore, an alternate hypothesis is also possible. Rotating between semantically unrelated stimuli may facilitate the generation of more innovative, original ideas. Lamm and Trommsdorff (1973) suggest that when stimuli are homogenous (as would be expected of other members' ideas in the maximum clustering condition) resultant ideas may be highly conventional because the accessible knowledge is restricted to a limited domain. By contrast, heterogeneous stimuli may facilitate "cross-fertilization," resulting in more innovative ideas (Nagasundaram & Dennis, 1993; Lamm & Trommsdorff, 1973). If there is a "cross-fertilization" effect of switching between categories that results from keeping more problem-relevant knowledge accessible, it may yield more original ideas. Thus, there is a plausible alternative hypothesis:

- *H_2: More high-quality ideas will be generated by those in the minimum clustering condition compared with those in the maximum clustering condition.*

Finally, according to Ward's (1994) *path-of-least-resistance* model, individuals

generate ideas using the least amount of cognitive effort possible. This explains the

observed tendency for people to start out by generating conventional ideas. It is not until

these ideas have been expressed that more original ideas are generated (Rietzschel et al.,

2007). If this model is correct, then, two additional outcomes are likely:

- *H_3: More highly-feasible ideas will be found among the first third of the ideas generated than among the last third of ideas generated.*

- *H_4: More highly-original ideas will be found among the last third of the ideas generated than among the first third of ideas generated.*

- *H_5: A higher proportion of high-quality ideas (using an index that combines feasibility and originality) will be found among the last two-thirds of ideas generated than among the first third of ideas generated.*

Nominal vs. Interacting Groups

Finally, cognitive stimulation and cognitive interference have been considered in terms of

the different impact they may have due to the amount of semantic clustering. However, it

is also probable that these two cognitive phenomena will operate differently in nominal

groups as opposed to interacting groups. Cognitive stimulation is one of the presumed

benefits of being exposed to the ideas of other group members (Nijstad & Stroebe, 2006).

These members' ideas may help activate knowledge that would not have been accessible

to the individual if he were working alone. Paulus and Yang (2000) compared nominal

groups to four-person interacting brainstorm groups using brainwriting. In that study, the

interacting groups generated 41% more ideas than did the nominal groups. Further, when

the interacting groups were later broken up to work as individuals, they still generated

almost twice the number of ideas as those who had been in the nominal groups originally.

This effect was thought to occur because brainwriting requires group members to pay

greater attention to the ideas of other members than does conventional brainstorming.

Greater attention to the ideas of others has been found to have beneficial cognitive

stimulation effects (Nijstad & Stroebe, 2006). There are two reasons to question whether

the results observed by Paulus and Yang (2000) will generalize to the current study.

First, the focus of that study was simply the quantity of ideas. Second, more recent

research has addressed a confound in the design of the Paulus and Yang (2000) study

related to the response format used in the different conditions and failed to replicate the

superiority of interacting groups (Goldenberg, Larson & Wiley, 2013).

There is also the possibility of cognitive interference in groups, which is the idea

that other members' ideas can disrupt one's own cognitive process of idea generation. In

the current study, because participants worked from the same semantic categories

provided by the experimenter, it is likely that members of the interacting brainwriting

groups worked from the same or a similar image. If so, this should reduce the amount of

cognitive interference (Nijstad et al., 2002). Therefore, cognitive interference is not

anticipated to strongly impact the interacting groups, whereas cognitive stimulation

should benefit the interacting groups.

Additionally, by working in groups, members may be forced to explore semantic

categories more deeply for another reason. In the nominal groups, those who will

eventually be grouped together as members of the same group will not see one another's

ideas. Therefore, there is a greater likelihood of the nominal groups generating more

duplicate ideas. However, the interacting groups will see one another's ideas. Members

will likely actively seek to avoid duplicating another member's idea. This avoidance

should force them to go deeper into the semantic category to find an idea not already

generated. Therefore, a final hypothesis is offered:

- *H_6: More high-quality ideas will be generated in the interacting*

  brainwriting groups than in the nominal brainwriting groups.

No predictions regarding interaction effects were made.

CHAPTER TWO

METHODS

Design and Participants

The study employed a two (clustering: minimum vs. maximum) by two (group: interacting vs. nominal) by two (semantic set: A vs. B) fully-crossed experimental design. For purposes of generalizability, two roughly comparable sets of semantic categories were used. One hundred and forty-eight Loyola University students participated in the study. For their involvement, they received research participation credit in their psychology course. Three participants' data were discarded due to a research assistant's error in running the brainwriting session. Data from four other participants from the nominal group condition were discarded because there were not enough other participants to form another complete group. The final sample consisted of 141 participants who formed 47 3-person groups approximately evenly distributed among the eight conditions (see Table 1).

Table 1. Number of group observations per experimental condition.

Condition

		Frequency	Percent	Valid Percent	Cumulative Percent
	SoloMinA	5	10.6	10.6	10.6
	SoloMinB	5	10.6	10.6	21.3
	SoloMaxA	6	12.8	12.8	34.0
	SoloMaxB	5	10.6	10.6	44.7
Valid	GroupMinA	7	14.9	14.9	59.6
	GroupMinB	6	12.8	12.8	72.3
	GroupMaxA	6	12.8	12.8	85.1
	GroupMaxB	7	14.9	14.9	100.0
	Total	47	100.0	100.0	

The amount of clustering, or average cluster length, was experimentally manipulated at two levels that will be referred to as minimum clustering and maximum clustering. The typical average cluster length for a participant in the minimum clustering condition is one. An idea stream[1] generated by an individual in this condition resembled the following pattern: $A_1 \rightarrow B_1 \rightarrow C_1 \rightarrow A_2 \rightarrow B_2 \rightarrow C_2 \rightarrow A_3 \rightarrow B_3 \rightarrow C_3 \rightarrow ... \rightarrow A_6 \rightarrow B_6 \rightarrow C_6$, yielding six ideas from each of 3 different semantic categories (for a total of 18 ideas per participant). The typical maximum clustering condition yielded an average cluster length of six, with an idea stream that resembled the following pattern: $A_1 A_2 A_{.3} ... A_6 \rightarrow B_1 B_2 B_3 ... B_6 \rightarrow C_1 C_2 C_3 ... C_6$. This also yields six ideas from each of 3 semantic categories. The key difference is that in the minimum clustering condition just

---

[1] Note that these are individuals' idea streams. Groups will consist of three members, so a group will generate 6 ideas for each of the three semantic categories resulting in a total of 18 ideas.

one idea was generated from a given semantic category before rotating to the next, whereas in the maximum clustering condition six ideas were generated from a given semantic category before switching to a new category. These idea streams are idealized; in reality some variation occurred. Participants sometimes failed to generate the desired number of ideas, or generated a repeat idea. Still, the observed cluster lengths in the maximum clustering conditions were substantially longer than the observed cluster lengths in the minimum clustering conditions.

<div align="center">

Procedures and Materials

</div>

Task and materials. Before their arrival, participants were randomly assigned to one of the experimental conditions. Nominal and interacting groups were run in separate sessions, but as many as 18 participants were run during the same session. The minimum and maximum clustering conditions were often run during the same session.

Upon arrival, participants were directed to a data collection room. In all conditions, instructions were given that include the following elements (see Appendix A): general description of the task, obtaining informed consent, an explanation of brainwriting, detailed explanations of how to correctly record ideas, a sample problem and an example of semantically-related ideas[2], and an opportunity for participants to ask questions. Only three of Osborn's four brainstorming rules were given. Because global productivity was held constant across conditions, the rule instructing participants to generate as many ideas as possible was excluded. After all participants understood the procedure, the session began and participants worked quietly.

---

[2] The example topic used during instructions was, "Ways to raise funds for a local charity." Example semantically-related categories given included "sales" and "events." Examples of ideas for the semantic-category "sales" included bake sales and used book sales.

Four-by-six inch index cards were labeled at the top with a semantic category prime intended to provoke ideas within that semantic category. The problem that was used was the "university problem," which has been used in a number of previous studies. Participants were asked to "Think of ways to improve Loyola University Chicago." One of two sets of semantic category primes that were intended to be comparable was assigned for use. Set A included: 1) teaching, 2) commuting, parking, and transportation, and 3) library. Set B included: 1) advising, 2) dorms/housing, and 3) bookstore.

Interacting groups. In the interacting group condition, participants were seated in a group of three participants around a table or a group of desks facing one another [see Figure 1]. Each participant had a deck of downward facing index cards and a different color pen (blue, black, or red) in front of them. The different color ink allowed the experimenter to differentiate between ideas from various group members during idea transcription. When instructed, each participant flipped over the top index card and read the semantic category written at the top. They each generated one idea within the given category, then passed the index card in a clockwise direction to another group member, and simultaneously received a card with one idea written on it from the other group member. The card received had either the same or a different semantic category written at the top, depending on the experimental condition. In the maximum clustering condition, each new card had the same semantic category written at the top until each member had generated six ideas from that category. In the minimum clustering condition, each new card had a different one of the three semantic categories at the top. The receiving member read the semantic category and the previously recorded idea, and

added his or her own idea within that semantic category. A new idea was recorded if at all possible; however, the option of recording an "X" to indicate the inability to generate a new idea was made available. This procedure continued until three ideas (or X's) were recorded on each card, at which point the card was placed face-down in the middle of the group. Then, a new card was flipped over by each member, and the process repeated. This continued until all cards had been used.
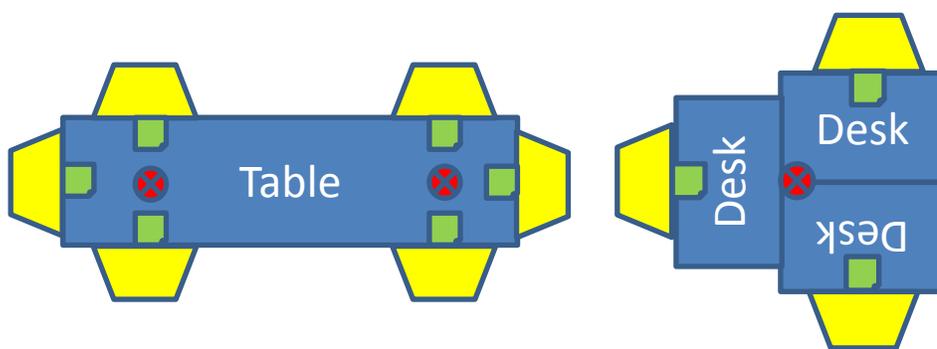

Figure 1. Seating arrangement in interacting condition.

Nominal groups. In nominal groups, multiple participants were seated individually and worked simultaneously [see Figure 2]. Each individual had his or her own set of index cards and a black, blue, or red pen. He or she flipped over an index card and read the semantic category written at the top. In the minimum clustering condition, the participants generated one idea for the given semantic category before flipping over a new card which had the next semantic category on it. Each new card had a different (of the three) semantic category at the top, and the three semantic categories appeared in rotation until each had been used by the participant to generate a total of 18 ideas. In the maximum clustering condition, the participant also generated one idea for the given semantic category before flipping over a new card; however the new card had the same

semantic category on it. So, each new card had the same semantic category on it until the participant had generated six ideas using that category. Thus, the nominal condition resembled the interacting group condition very closely except that participants did not pass or receive cards from any other participants, and only one idea was recorded per note card.
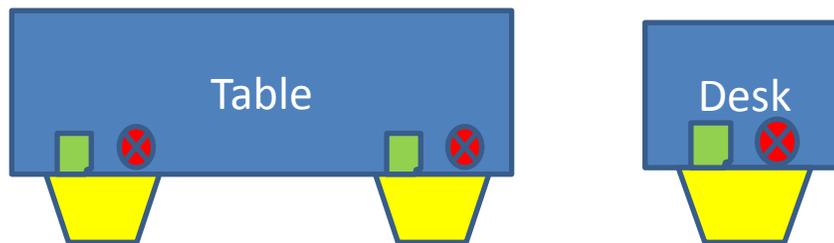

Figure 2. Seating arrangement in individual condition.

Post-session. After all participants finished with idea generation, a post-session questionnaire was filled out by each participant (see Appendix B). The questionnaire contained items assessing comprehension of the task and clarity of the instruction. Also, the questionnaire included items to determine whether various experimenters acted consistently in running the sessions. These questions asked how enthusiastic the experimenter appeared to be and how important participants felt it was to the experimenter that they do their best on the task. It also assessed task enjoyment and perceptions of one's performance on the task, including the amount of cognitive failure experienced. The questionnaire also asked how many additional ideas the participant believed that he could have generated within each category. After all participants finished the post-session questionnaire, they were debriefed, thanked, and dismissed. The session took about thirty minutes on average.

Dependent Measures

A broad but generally accepted definition of quality that was used in conjunction with the above definition states that it is "some combination of *originality* (the degree to which an idea is innovative) and *appropriateness* (e.g., the degree to which a product or an idea is relevant to the topic, or is thought to be practically feasible)" (italics in original) (Rietzschel, Nijstad & Stroebe, 2007). Working from that definition, the precise definition of a "high-quality idea" used in this study was, "a phrase that was proposed as a solution for the problem at hand that was both highly-original, and highly-feasible." Originality and feasibility were assessed using measures similar to those used by Rietzschel et al. (2007). Ideas were scored on two six point scales (ranging from 1 = "not at all original/feasible" to 6 = "highly original/feasible"). Scores were derived for originality and feasibility of the ideas compared with other ideas generated from the same semantic prompt using a Q-sort. Originality should be thought of as being novel or creative. Feasible should be thought of as being "relevant to the problem and implementable given available resources (e.g., space, funds) which do not generate additional problems."

The originality and feasibility ratings were considered together to determine whether or not an idea qualified as a high-quality idea. As discussed in the introduction, a good-idea-count was chosen as the best measure (Paulus & Brown, 2002) of feasibility, originality, and overall quality. In the current study, a highly-original idea was one that received an originality score greater than four, and a highly-feasible idea was one that

received a feasibility score greater than four. A high-quality idea was one that received above a four on both subscales.

Content Coding

After ideas were collected, the data was coded in multiple phases. I transcribed all ideas and assigned a unique identifier to each one. The code associated the idea with the group type, clustering condition, and the third of the brainwriting session in which the idea was generated. Raters did not know the meaning of the digits in the identifier; which allowed the raters to remain blind to the condition in which each idea was generated. Then all ideas were printed onto labels, transferred to index cards, and sorted into six decks based on their semantic prompt.

Four undergraduate research assistants received training that covered a conceptual understanding of the constructs feasibility and originality, and also a procedural understanding of how to perform a Q-sort. Scores were obtained using a multi-step process. First, raters sorted the ideas in each deck into three categories – high, medium, and low (feasibility or originality, respectively).   The raters were instructed to put approximately 40% of the ideas in the medium category with 30% each in the high and low categories. Next, they sorted the ideas in each category into two sub-categories – high and low. Again the raters were given specific instructions about the percent of ideas that should be assigned to each category. Specifically, the high and low categories were divided so that a smaller percentage of the ideas were placed in the more extreme categories. This process ensured that the ratings resulted in six category assignments that were approximately normally distributed.

Percent agreement was used as a measure of interrater reliability. Percent agreement was calculated for each pair of raters, at the group level. Recall that the dependent variable used for each of the measures is good-idea count and that to qualify as a good-idea an idea must meet the cut-off of greater than four on its respective scale. Percent agreement in this case then refers to the degree to which the raters agreed on the number of highly-original or highly-feasible ideas generated by a given group. For feasibility ratings, percent agreement ranged from 82.75% to 93.24% with an average of 86.32%. For originality ratings, percent agreement ranged from 76.24% to 93.86% with an average of 82.79% (see Table 2.) In general, agreement was high suggesting good reliability between raters. The final feasibility/originality score used to determine good-idea count was the average of the two scores assigned by the two different raters.

Table 2. Interrater reliability expressed as percent agreement.

Semantic Set	Semantic Category	Percent Agreement	
		Feasibility	Originality
A	Commuting, Parking, & Transport.	83.41%	78.13%
	Library	82.75%	76.24%
	Teaching	84.62%	84.46%
B	Advising	88.45%	93.86%
	Bookstore	93.24%	85.52%
	Dorms/Housing	85.43%	76.24%

After feasibility and originality ratings were determined for each idea, nominal groups were formed. A nominal group is a group in name only. In reality, the three participants worked individually, but their ideas were considered as part of a group for the purpose of analysis. Specifically, after ideas were collected and transcribed, they were put in chronological order by the date of the session. The first three sets of ideas

collected from the same condition formed the first nominal group. The next three sets of ideas collected formed the second nominal group, and so on until all sets of ideas have been assembled into groups of three.

The cut-off of greater than four was used to determine the good idea count of each of the three dependent variables – number of highly-feasible ideas, number of highly-original ideas, and number of high-quality ideas – generated by participants in each condition.

Finally, duplicate ideas were removed from ideas generated in the nominal group condition. The idea cards were divided by semantic category and sorted into the nominal groups to which they were assigned. Those sets of cards were given to an undergraduate research assistant who was blind to the conditions and hypotheses of the study. He compared the cards to one another and determined which ideas were duplicates. Whenever the feasibility or originality rating of two ideas determined to be duplicates varied from one another, the average of the two ratings was used.

CHAPTER THREE

RESULTS

Preliminary Analysis

The inclusion of two different semantic sets was intended to minimize the chance that

any significant findings were an anomaly related to the specific semantic prompts used.

As such, the hope was that there would be no significant differences related to the

semantic set used. A two (group type: nominal vs. interacting) by two (clustering:

minimum vs. maximum) by two (semantic set: A vs. B) MANOVA confirmed that there

were no significant mean difference or interaction effects of the semantic set used on any

of the dependent variables: number of highly-feasible ideas $F(1, 46)=.03$, $p = .86$; number

of highly-original ideas $F(1, 46) = .02$, $p = .90$; number of high-quality ideas $F(1, 46) =$

1.10, $p = .30$; or cognitive failure $F (1, 46) = .01$, $p = .93$.

Similarly, a separate two (group type: nominal vs. interacting) by two (clustering:

minimum vs. maximum) by two (semantic set: A vs. B) MANOVA performed on the

data collected from the post-session questionnaire corroborate these findings.

Specifically, based on the semantic set used, participants did not report differences in task

enjoyment (A: $M = 2.11$, B: $M = 2.15$) $F(1, 150) = .05$, $p = .83$, perceived performance

(A: $M = 2.85$, B: $M = 2.84$) $F(1, 150) = .00$, $p = .96$, task difficulty (A: $M = 2.45$, B: $M =$

2.77) $F(1, 150) = 3.21$, $p = .08$, self-reported cognitive failure (A: $M = 1.27$, B: $M =$

1.41), $F(1, 150) = .60$, $p = .44$, or the frequency with which previously generated ideas

reoccurred to them (A: $M = 2.77$, B: $M = 2.67$), $F(1, 150) = .26$, $p = .61$. Nor were there

any significant interaction effects involving the semantic set with any of the other

independent variables based on the post-session questionnaire data. Thus, the

independent variable semantic set was collapsed across for the remainder of the analysis.

While no specific predictions were made regarding the amount of cognitive

failure experienced in various conditions, there was an underlying assumption that more

cognitive failure would be experienced in the maximum clustering condition and by

individuals as opposed to groups. This assumption played a role in the development of

some of the hypotheses. As a check of this expectation, a 2 (clustering: minimum vs.

maximum) x 2 (group: nominal vs. interacting) ANOVA was conducted. This analysis

confirmed that individuals working alone ($M = 4.81$) experienced more cognitive failure

than those working in interacting groups ($M = 2.04$), $F (1. 46) = 7.35$, $p = .01$. However,

there was not a significant difference in the amount of cognitive failure experienced by

those in the minimum clustering condition ($M = 3.78$) compared with those in the

maximum clustering condition ($M = 2.79$), $F (1, 46) = 1.06$, $p = .31$.

Hypotheses Tests

Hypothesis 1 and 2 make opposite predictions about the quality of the ideas generated

using the minimum versus maximum clustering paradigm. Specifically, Hypothesis 1

states that more high-quality ideas will be generated by those in the maximum clustering

condition than by those in the minimum clustering condition; whereas, Hypothesis 2

states that more high-quality ideas will be generated by those in the minimum clustering

condition than by those in the maximum clustering condition. They share a single null

hypothesis, that there will be no significant mean difference in the number of high-quality

ideas generated using the two different clustering paradigms. Therefore, a single 2

(clustering: minimum vs. maximum) x 2 (group: nominal vs. interacting) ANOVA was

used to test these opposing hypotheses. This analysis revealed that there was not a

significantly different number of high quality ideas generated by those in the minimum

clustering condition ($M = 5.48$) and the maximum clustering condition ($M = 5.19$), $F(1,$

$46) = .11$, $p = .75$ [Figure 3]. Thus we failed to reject the shared null hypothesis for
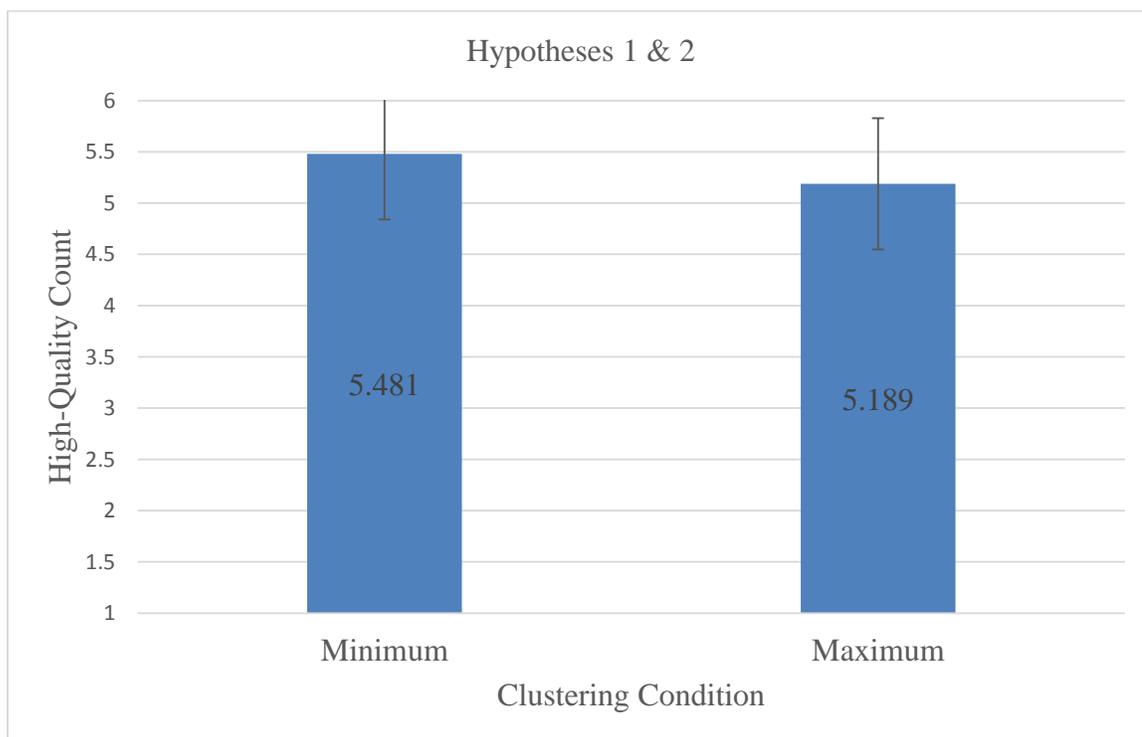
Hypotheses 1 and 2.

Figure 3. Results of Hypotheses 1 and 2 tests.

The same 2 (clustering: minimum vs. maximum) x 2 (group: nominal vs.

interacting) ANOVA was used to test Hypothesis 6 that states that more high-quality

ideas will be generated in the interacting brainwriting groups than in the nominal

brainwriting groups. It revealed that while those in interacting groups generated slightly more high-quality ideas ($M = 5.692$) than those working alone ($M = 4.977$), the mean between-group difference was not significant, $F(1, 46) = .63$, $p = .43$ [Figure 4.] Thus we failed to reject the null hypothesis for Hypothesis 6.
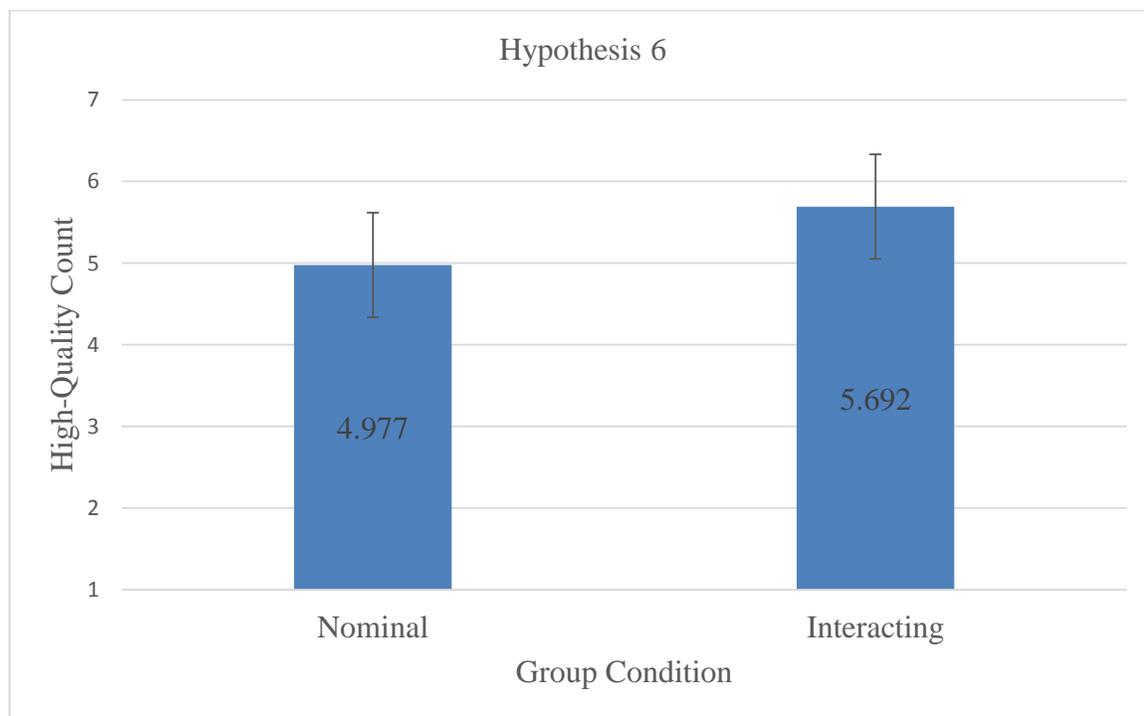

Figure 4. Results of Hypothesis 6 test.

A series of paired samples t-tests was used to test Hypotheses 3 and 4. Hypothesis 3 states that more highly-feasible ideas will be found among the first third of the ideas generated than among the last third of the ideas generated. The paired samples t-test revealed that, indeed, significantly more highly feasible ideas were generated during the first third ($M = 5.47$) of the brainwriting session compared with the last third ($M = 4.70$), $t(47) = 2.31$, $p = .03$ [Figure 5.] Thus the null hypothesis is rejected, and Hypothesis 3 is accepted.
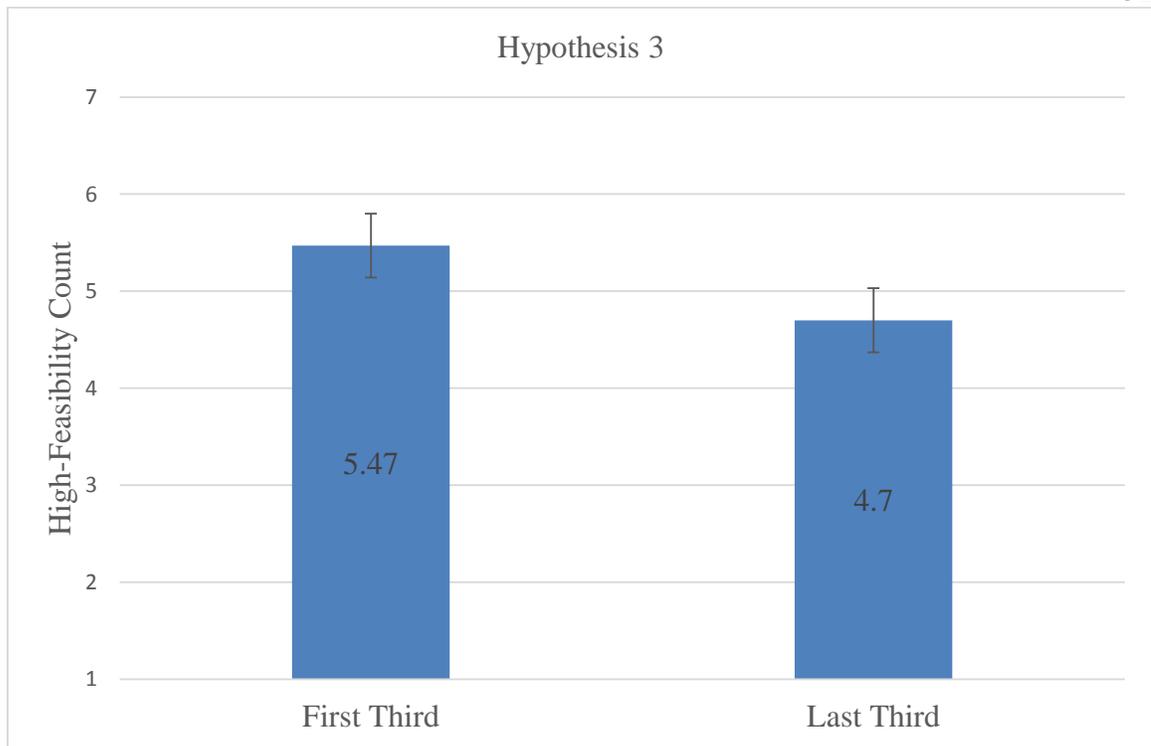
Figure 5. Results of Hypothesis 3 test.

Hypothesis 4 states that more highly-original ideas will be found among the last third of ideas generated than among the first third of ideas generated. The paired samples t-test confirmed that significantly more highly original ideas were indeed generated during the last third ($M = 5.49$) of the brainwriting session compared with the first third ($M = 4.55$) of the brainwriting session, $t(47) = -2.64$, $p = .01$ [Figure 6.] Thus the null hypothesis was rejected, and Hypothesis 4 was accepted.
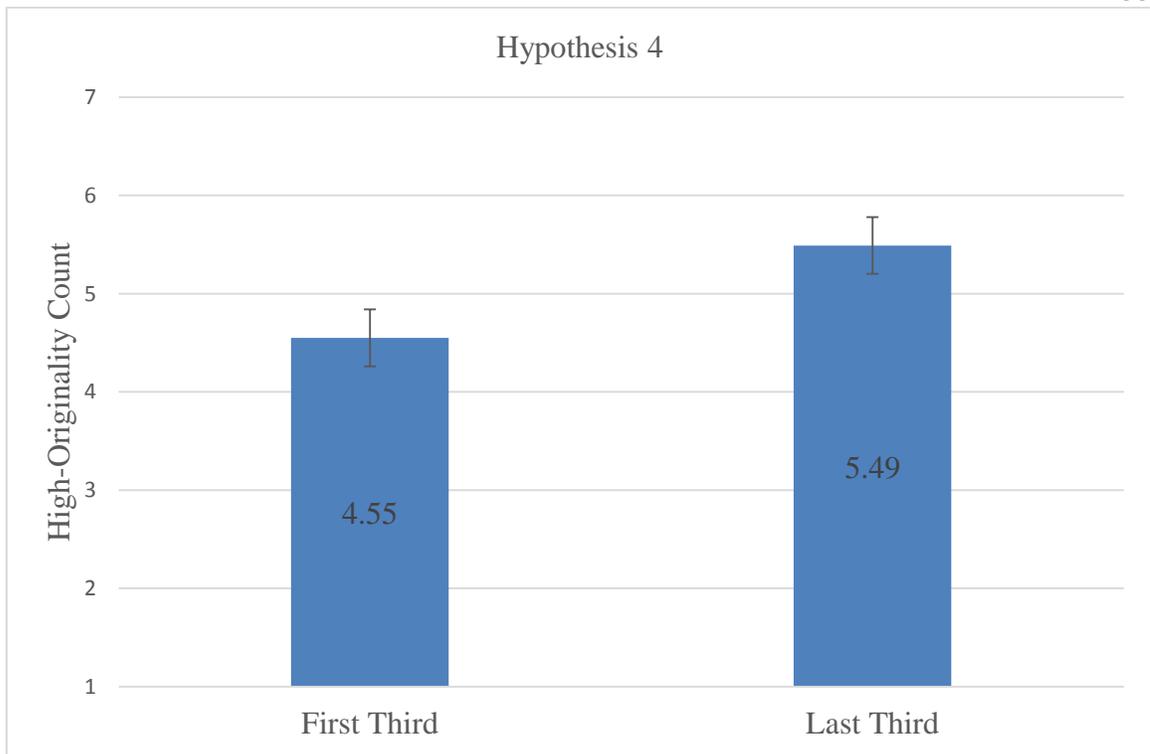
Figure 6. Results of hypothesis 4 test.

Hypothesis 5 states that a higher proportion of high-quality ideas (using an index that combines feasibility and originality) will be found among the last two-thirds of the brainwriting session than during the first third of the session. A two (thirds: first third vs. last two-thirds; within) by two (clustering: minimum vs. maximum; between) by two (group: nominal vs. interacting; between) mixed model ANOVA was run to test this hypothesis. It revealed that there was not a significant difference in the percentage of high-quality ideas generated during the first third ($M = 12.48\%$) versus the last two-thirds ($M = 10.69\%$) of the brainwriting session, $F (1, 43) = 1.21, p = .28$ [Figure 7.] Thus we failed to reject the null hypothesis for Hypothesis 5.
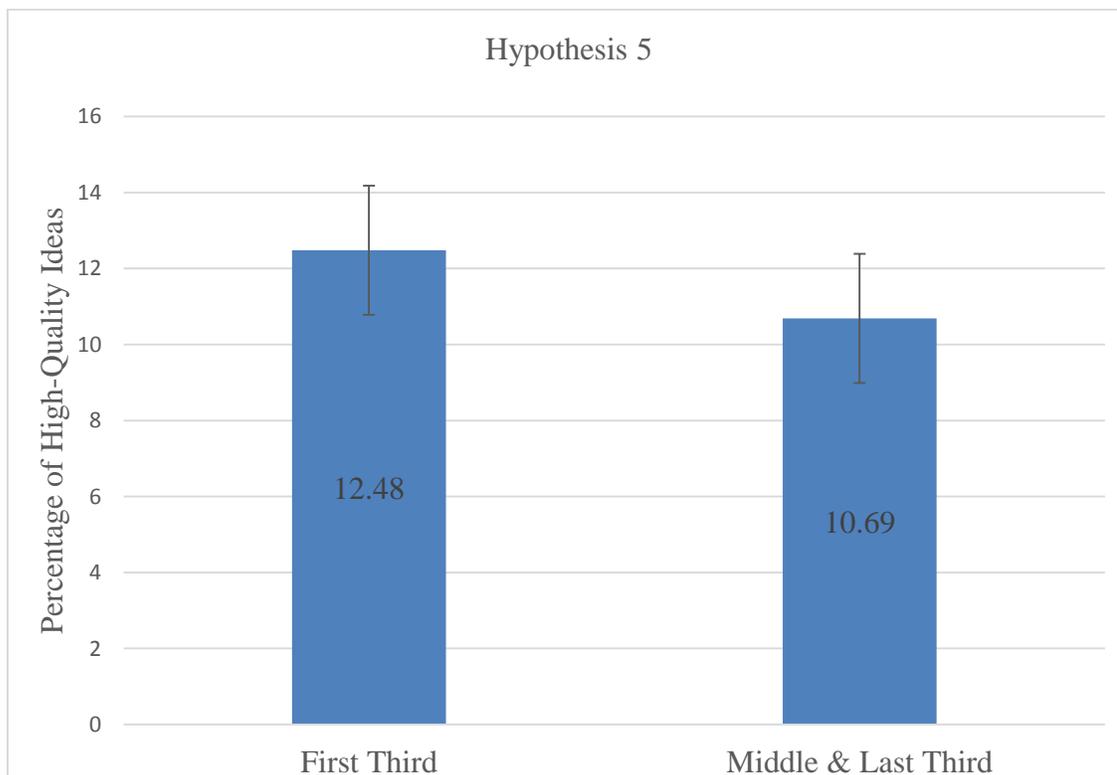
Figure 7. Results of hypothesis 5 test.

Exploratory Analysis

While no specific predictions were made with regards to interaction effects, the mixed

model ANOVA used to test Hypothesis 5 revealed a marginally significant thirds (first

third vs. last two-thirds) by clustering interaction, $F(1, 43) = 4.04$, $p = .05$[1] [Figure 8.]

To understand the nature of this significant interaction, the file was split by clustering

condition and a paired samples t-test was run comparing the percentage of high-quality

ideas generated during the first third versus the last two-thirds of the brainwriting session.

This analysis revealed that for those in the minimum clustering condition, the percentage

of high-quality ideas generated during the first third ($M = 10.78\%$) was not significant

different from the percentage of high-quality ideas generated during the last two-thirds

---

[1] $p = .05063$

($M = 12.52\%$), $t(22) = -.75$, $p = .46$. However, those in the maximum clustering

condition generated significantly more high-quality ideas during the first third ($M =$

14.12%) than during the last two-thirds ($M = 8.94\%$), $t(23) = 2.23$, $p = .04$. This pattern

of results suggests that while those in the minimum clustering condition started out by

generating a slightly lower percentage of high-quality ideas than did those in the

maximum clustering condition, their ability to continue generating high-quality ideas did

not decrease significantly over the course of the brainwriting session while that of those

in the maximum clustering condition did. It is possible that, had participants been asked

to generate more ideas, this trend of diminished ability to generate high-quality ideas by

those using maximum clustering would have resulted in a significant overall main effect
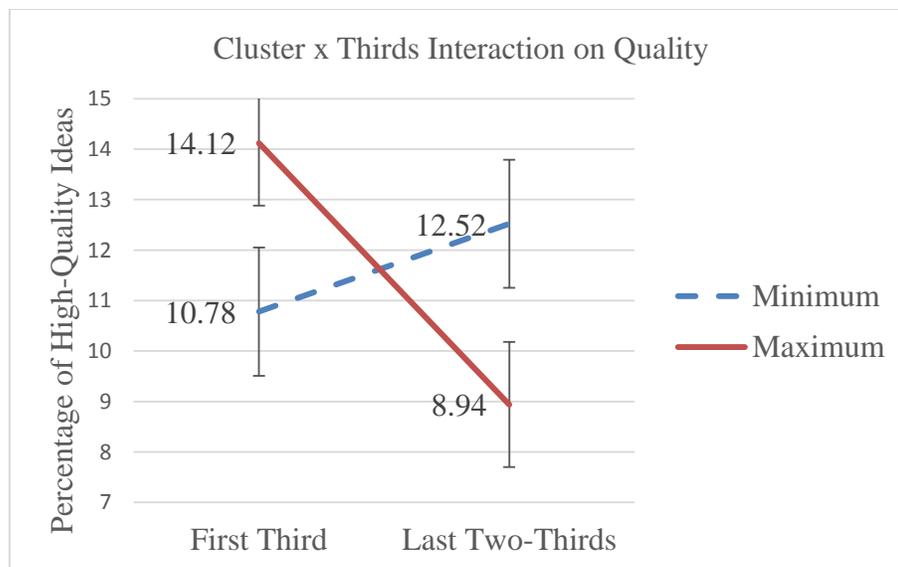
of the clustering manipulation.

Figure 8. Marginally significant cluster by thirds interaction on idea quality.

Based on this significant interaction, I decided to test whether there was a significant thirds-by-clustering interaction for the dependent measure originality[2]. A 2 (thirds: first vs. last two-thirds; within) by 2 (clustering: minimum vs. maximum; between) by 2 (group: nominal vs. interacting; between) mixed model ANOVA revealed a significant main effect of thirds such that there was a significantly higher percentage of highly-original ideas generated during the last two-thirds of the brainwriting session ($M = 34.63\%$) compared with the first third ($M = 27.78\%$), $F(1, 43) = 12.68$, $p = .001$[3]. There was also a significant thirds-by-clustering interaction, $F(1, 43) = 11.34$, $p = .002$ [Figure 9.] Again, the data set was split by cluster condition and a paired samples t-test was conducted to determine the nature of the significant interaction. This analysis revealed that those in the maximum clustering condition generated a similar proportion of highly-original ideas during the first third ($M = 30.28\%$) compared to the last two-thirds ($M = 30.91\%$) of the brainwriting session, $t(23) = -.23$, $p = .82$. However, those in the minimum clustering condition generated a significantly higher proportion of high-quality ideas during the last two-thirds ($M = 38.51\%$) compared with the first third ($M = 25.17\%$) of the brainwriting session, $t(22) = -4.71$, $p = .00$. So, while there was not a significant main effect of cluster type, $F(1, 43) = .15$, $p = .70$, those in the minimum clustering condition showed a significant increase in the percentage of highly-original ideas that they generated over the course of the brainwriting session.

---

[2] I also conducted a mixed model ANOVA to look for a significant thirds-by-clustering interaction effect for the DV feasibility, but there were no significant results to discuss. Possible reasons for this lack of a similar interaction effect for feasibility are explored in the discussion section.

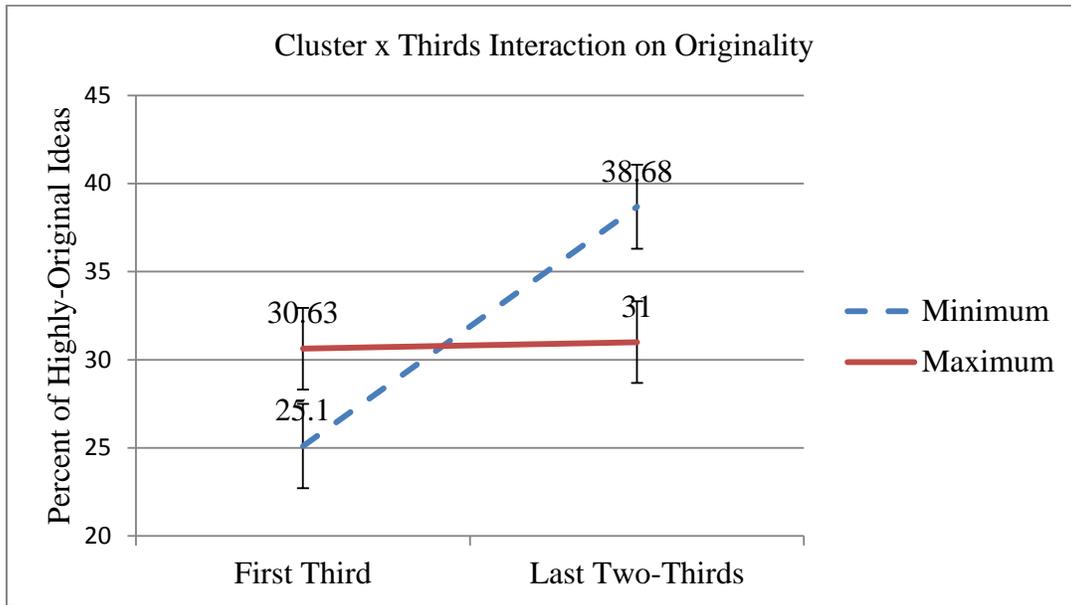[3] Which is consistent with the rejection of the null hypothesis for Hypothesis 4.

Figure 9. Significant cluster by thirds interaction on idea quality.

CHAPTER FOUR

DISCUSSION

Summary

The primary goal of this study was to understand whether the benefits of deep exploration within a semantic category for idea quality depend on the way in which the semantic category is explored (Rietzschel, Nijstad & Stroebe, 2007). In particular, this study attempted to unravel whether it is fluency in general or clustering in particular that mattered. The pattern of results observed in the current study suggest that fluency is primarily responsible for this effect, but also that over the course of a brainwriting session a minimum clustering approach yields a higher proportion of highly-original ideas and is better suited to sustaining the quality of the ideas generated. While there was not a significant main effect of clustering on any of the DVs, there was a significant thirds-by-clustering interaction on both the number of highly-original ideas and the number of high-quality ideas that participants generated.

There was not a significant main effect of group type, which is not altogether surprising. While individuals have historically outperformed interacting groups in terms of global productivity, the findings have been mixed when the dependent variable is idea quality, often finding the performance of individuals and groups to be similar. What is somewhat more interesting is that the direction of the effects suggests a cross-over

interaction between group type and clustering condition, although this interaction was not significant [see Figure 10]. The pattern indicates that interacting groups performed better in the minimum clustering condition whereas individuals performed better in the maximum clustering condition. Interestingly, if this is the case, it suggests possibly that cognitive stimulation from other group members caused the minimum clustering paradigm to be beneficial to interacting groups, whereas individuals who did not have the benefit of cognitive stimulation for some reason found the minimum clustering condition to be less conducive to generating high-quality ideas. Thus, cognitive stimulation may have played a role. This suggestion should be interpreted with great caution, however, as the between group differences for interacting and nominal groups was not significant.
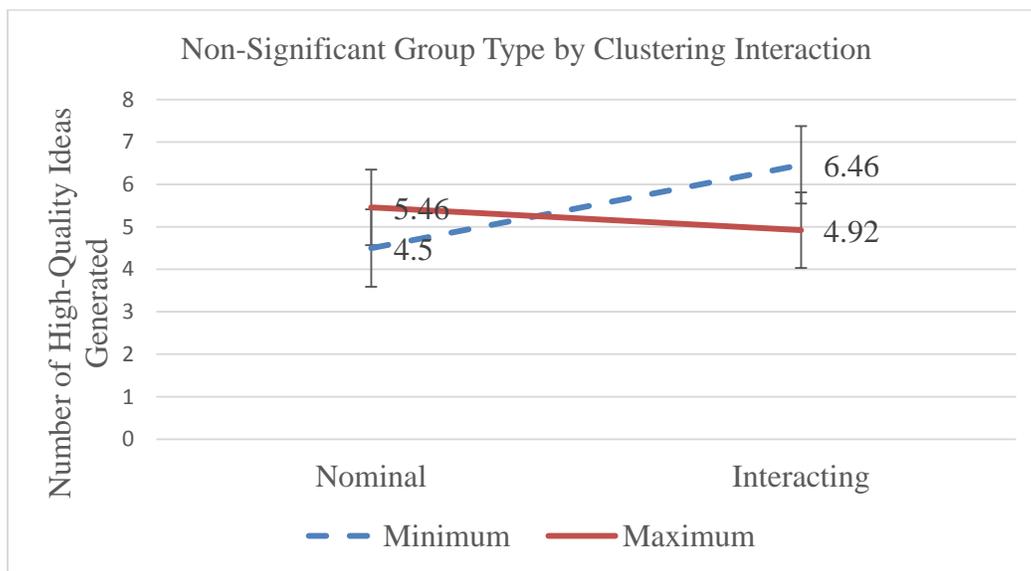

Figure 10. Non-significant group type by clustering interaction on the number of good-ideas generated.

The absence of a significant main effect of the clustering manipulation is more surprising and could be due to a variety of causes. At a minimum, it indicates that

rotating between three semantic categories did not impair the performance of those in the minimum clustering condition. For the SIAM model, this either means that participants were able to switch easily between the images being used, or that they were able to work simultaneously from multiple images. A bolder assertion would be that this lack of a significant main effect suggests that it is depth of exploration in general and not the specific way in which a category is explored that matters. The current findings do not rule out this possibility. However, the significant thirds-by-clustering interaction on originality and quality suggest a more nuanced explanation.

It is conceivable this pattern of results has to do with the order in which semantic categories were presented to participants. Semantic Set A was always presented in the following order: (1) Library, (2) Teaching, and (3) Commuting, Parking, and Transportation. Semantic Set B was always presented in the following order: (1) Bookstore, (2) Advising, and (3) Dorms and Housing. Thus, a participant in the maximum clustering condition will always have generated the first third of their ideas using either Library or Bookstore, the second third using Teaching or Advising, and the last third using Commuting, Parking and Transportation or Dorms and Housing – depending on their semantic set condition. It follows then that if Library and/or Bookstore were more difficult semantic categories that the performance of those in the maximum clustering condition during the first third of the brainwriting session would appear to be inferior not as a result of the clustering manipulation but because of the difficulty of the semantic subcategory. However, this explanation seems unlikely for two reasons.

First, raters rated the originality and feasibility of ideas in relation to others within the same semantic category. In other words, library-related ideas were always compared to other library-related ideas (never to teaching-related ideas). The Q-sort allowed the most feasible and original ideas generated within each semantic category to be identified, minimizing inherent differences between the semantic categories used. Second, the absence of significant between-group differences on any of the dependent variables also suggests that the effects were not related to inherent differences in the semantic categories used. Thus, the more likely explanation is that there is something about using minimum or maximum clustering that produces actual differences in the participants' performance.

Turning now to the significant interaction effects, those in the maximum clustering condition tended to start out by generating a slightly higher proportion of high-quality and highly-original ideas than did those in the minimum clustering condition. However, by the end of the brainwriting session, the pattern had reversed for both of those dependent variables. The percentage of high-quality ideas generated by those in the maximum clustering condition was significantly lower during the last two-thirds of the brainwriting session compared with the first third, while the percentage of high-quality ideas generated by those in the minimum clustering condition did not drop, and in fact increased slightly over the course of the brainwriting session. The percentage of highly-original ideas generated by those in the maximum clustering condition changed very little over the course of the brainwriting session, whereas the percentage of highly-original

ideas generated by those in the minimum clustering condition increased significantly from the first-third to the last two-thirds.

So, there are two significant interaction effects of interest. First, the percentage of high-quality ideas generated by those in the maximum clustering condition dropped significantly during the last two-thirds of the idea generation session. Second, the percentage of highly-original ideas generated by those in the minimum clustering condition increased significantly during the last two-thirds of the idea generation session. Next, I will explore the implications and possible causes of these two effects.

The significant decrease in the proportion of high-quality ideas among those in the maximum clustering condition can be understood from a motivation perspective. Perhaps participants were demotivated because the task was harder, less stimulating, or less enjoyable than in the minimum clustering condition. If it were harder, one might expect to see more cognitive failure in the maximum clustering condition, but that was not the case. If it were less stimulating or enjoyable, one might expect to see lower self-reported levels of task enjoyment, or higher perceptions of task difficulty, on the post-session questionnaire. Yet none of this was true either. Thus, it seems that, unbeknownst to the participants, there was a true difference between minimum and maximum clustering that resulted in a higher proportion of high-quality ideas found among later ideas generated in the minimum clustering condition. While this rules out a motivational explanation, it does not speak to the cause of the difference induced by the clustering manipulation. In an attempt to provide such an explanation, and because high-quality

ideas were determined based on the feasibility and originality of the ideas, I will now turn to the second significant interaction, related to originality.

The significant thirds-by-clustering interaction on originality can be taken as evidence of "cross-pollination" between semantic categories that resulted in more highly-original and, consequently, more high-quality ideas. This has important implications for the SIAM model. Specifically, the SIAM model theorizes that only one image is worked from at a time during the idea-generation loop, and that the current image is replaced by a new one anytime one returns to the image-retrieval loop. However, the significant increase in the percentage of highly-original ideas generated over the course of the brainwriting session by those only in the minimum clustering condition suggests that participants may have been able to use problem-relevant information activated by the first semantic category when generating the next idea from the second and perhaps third semantic category. So, perhaps not all of the problem-relevant information activated by one image is lost when a new image is retrieved.

Taken together, these two explanations provide the following rationale for the two significant thirds-by-clustering interactions. Participants in the minimum clustering condition benefitted from "cross-pollination" between problem-relevant knowledge kept active by rotating between the three semantic subcategories. Thus, they were able to generate a significantly greater proportion of highly-original ideas during the last two-thirds of the idea generation session. Because quality partially depended on originality, those highly-original ideas that were also highly-feasible were high-quality ideas. This translated into the second significant interaction effect; the significant difference in the

proportion of high-quality ideas generated between the clustering conditions during the last two-thirds of the idea generation session. Essentially, those in the maximum clustering condition did not benefit from cross-pollination. So, while the feasibility of their ideas naturally tended to diminished over the course of the session, and because they did not experience the benefit of cross-pollination on the originality of their ideas, the ideas that they generated toward the end of the session tended to be low-quality ideas. On the other hand, those in the minimum clustering condition did benefit from cross-pollination between subcategories and were thus able to continue to generate high-quality ideas toward the end of the session.

The lack of a similar, significant clustering-by-thirds interaction for the proportion of highly-feasible ideas could have something to do with the fact that this comparison was made between the first two-thirds and the last third (rather than the first third and the last two-thirds, as it was for both originality and quality). Perhaps a sufficient number of highly-feasible ideas existed in the solution space for the "University Problem" so that a longer brainwriting session would have been required in order to exhaust the highly-feasible ideas available. A certain degree of variability was imposed by the Q-sort procedure, in that ideas were rated based on their feasibility in relation to other ideas generated in this experiment. This allowed raters to make a distinction between the most and least feasible ideas even if all of them were relatively feasible. However, it is possible that the actual amount of variance represented in the feasibility of the ideas generated was not great enough to find meaningful differences on this dependent variables.

Practical Implications

The current study may help to inform best practices for the use of brainstorming. For instance, especially when dealing with interacting idea-generation groups, practitioners may want to choose carefully a few semantic subcategories that are thought to be particularly fertile ground for idea-generation. The subcategories should all relate to the same problem, but be different enough that they will activate somewhat different pieces of problem-relevant knowledge. These categories could then be presented in a fashion that promotes rotating through the categories during the idea-generation session. For example, as was done in this study, index cards or sheets of paper with the semantic category written at the top could be passed in round-robin fashion such that participants use the categories in rotating order. Alternatively, this could be achieved chronologically wherein each semantic category is employed for a period of seconds or minutes before proceeding to the next semantic subcategory. Eventually, the group would rotate back to the initial category and the process would repeat. These procedures both effectively promote a minimum clustering approach to idea generation. The current study also seems to illustrate the need to allow a brainstorm session to continue long enough to exhaust the more feasible ideas.

Limitations

The greatest limitation to the current study is the relatively small number of ideas generated by the brainwriting groups. Quantity was intentionally held constant at eighteen ideas per participant in order to see the unique effect of clustering on quality regardless of fluency or global productivity. However, if the marginally significant

thirds-by-clustering interaction found on the percentage of high-quality and highly-original ideas is a true indicator of how the two clustering paradigms would have operated over time, then eighteen ideas may have been too few to observe significant main effects of clustering. Thus, it would be preferable to repeat the current study having groups generate more ideas per topic or to generate ideas for more than three topics.

The decision to use three semantic subcategories was partially arbitrary, but also partially based on information we had regarding the "University Problem" used in this study. While choosing semantic categories for use, we referred to previous research (Goldenberg, Larson & Wiley, 2013) that had employed this problem. Working from the ideas generated, those authors identified 22 distinct semantic categories. From that list, semantic categories for the current study were selected to meet two goals. The first was to choose topics that were different enough so that ideas generated using one semantic prime were not likely to be similar to ideas generated using a different semantic prime. In other words, I wanted to avoid tapping into similar underlying images with different semantic primes in order to keep the depth obtained in each semantic category roughly equal. The second goal was to create two comparable sets of semantic categories that were approximately of equal difficulty. Two sets of three semantic categories each was selected that best met these goals.

There is a possibility that the Q-sort interfered with the ability to observe effects in the maximum clustering condition because the third during which an idea was generated was confounded with the semantic category. The Q-sort ensured that an approximately equal number of ideas from each semantic category would meet the high-

originality/high-feasibility cut-off. Thus, for those in the maximum clustering condition, it made it more likely that an equal number of ideas that they generated during each third would be rated as highly-original/highly-feasible. This may have caused the performance of participants in the maximum clustering condition to appear more similar across thirds than it actually was. Because there was a significant thirds-by-clustering interaction wherein those in the maximum clustering condition generated a lower percentage of high-quality ideas during the last two-thirds compared with the first third of the brainwriting session, it would appear that this limitation did not completely prohibit the ability to observe significant differences.

A final limitation of this study, which is common to research on idea generation, is the subjectivity inherent in conceptualizing and measuring quality as a construct. The operationalization of quality in this study gave equal weight to the feasibility and originality of the idea. However, it is possible that in a real world setting, the demands of the task prescribe greater weight to one construct or the other. It is also possible that feasibility and originality alone fail to fully reflect what is meant by a high-quality idea. There may be additional features, such as likelihood that the idea will be well received, that impact the desirability of a given idea in a real-world environment.

Future Directions

As mentioned above, the current study suggests that working from multiple semantic categories did not impair, and may have aided, generating highly-original and high-quality ideas. While three semantic subcategories were used in the current study, it is likely that there is a minimum, maximum, and optimal number of images that can be

simultaneously employ in idea generation. The actual optimal number likely depends on a variety of features unique to the individual, problem, and situational factors of the idea generation session. Individual variables likely to play a role include an idea generators' intellectual ability, motivation, familiarity with problem-relevant information, need for cognition, need for closure, and executive functioning skills. Important features of the problem may include the solution space, similarity or differences in the semantic categories, and goals of the idea-generation task (quantity vs. quality, ingenuity vs. feasibility, etc.) Finally, situational factors that may be relevant include the type of brainstorming session (EBS vs. verbal), group type, and group size. Thus, future studies may aim to address these questions directly. For instance, it may be of value to explore how the degree to which semantic categories are similar versus different impacts the ease with which ideas are generated using those categories, as well as the type of ideas that are generated using them in conjunction. In the current study, the topics represented within a semantic set were explicitly chosen to be dissimilar to one another with the aim of avoiding overlap between categories to attempt to keep depth of exploration roughly equal among them. Hypothetically, two similar or complementary semantic categories will activate similar or related problem-relevant information (respectively) and may encourage greater depth of exploration. Conversely, two very dissimilar semantic categories will likely elicit the use of two different images with two sets of problem-relevant information that overlap to a lesser extent. Having dissimilar problem-relevant information activated simultaneously may have different benefits such encouraging "out-

of-the-box" thinking. Indeed, the current study suggests the possibility of cross-

pollination between semantic subcategories. Further research in this area is encouraged.

APPENDIX A

EXPERIMENTER INSTRUCTIONS FOR

RUNNING A DATA COLLECTION SESSION

Before Session

    A. Print list of all expected Participants (Ps).

    B. Assemble all materials

        a. Sets of notecards prepared for session

            i. Labeled with semantic categories

            ii. Numbered on left 1-3

            iii. Numbered in bottom right-hand corner with member number and sequence number (e.g. C3 = member C's 3rd card)

            iv. Stacks rubber banded together. Enough for expected Ps plus extra.

        b. Pens: blue, black, and red. Be sure to have extras.

        c. Rubber bands

        d. Session labeling slips

        e. Large Zip-lock bags

        f. Post-it notes

        g. Experimenter (E) instructions

        h. List of Ps in order that they appear on Experimetrix

        i. Informed consent forms

        j. Debriefing letters

        k. Participant waiting area sign

        l. Late participant sign.

Beginning of Session

    A. Arrive 30 minutes prior to session.

        a. Hang Participant waiting area sign

        b. Arrange chairs/tables in data collection room

        c. Organize materials (i.e., notecards, consent forms, etc.)

    B. Greet Ps in meeting area

        a. Wait 5 minutes past session start time for any late-comers. Then exchange participant waiting area sign for late participant sign.

        b. Verify that al Ps are there for correct experiment

        c. Ask all Ps to turn off cell phones and store ALL personal items in bags

    C. In data collection room

        a. Have Ps place bags against one wall

        b. Have Ps check-in with you (this is so that you can be sure to assign credit to all those who showed up).

     c. Have Ps be seated in order that they appear on Experimetrix.

Instructions

  A. Introduction
        a. No talking with other Ps during any part of today's session: *"Please, no talking with any other participants during any part of today's session."*
        b. General description of task: "*During today's session, you will be given a topic and asked to generate a specific number of ideas or solutions. Ideas will be written on notecards. After you've finished generating ideas, you will be given a brief post-session questionnaire to complete. Everything will take less than an hour, and you will be assigned one experiment credit for your participation. Now that you know more about today's session, I'll pass around informed consent forms. Please feel free to take a moment to look over those, then sign and date them at the bottom. I'll be collecting them momentarily. If you'd like a copy of the informed consent document to take with you today, please just ask for one on your way out after the session today."*
        c. Informed Consent: E passes out and collects forms.
  B. Task
        a. Brainwriting Task
            i. Brainstorming: "*Brainstorming is a technique to help people better generate ideas or solutions to a problem. There are three guiding principles to brainstorming that you follow as you complete today's task: 1) avoid being critical of any idea, 2) feel free to combine ideas or build-upon ideas, and 3) wild or unusual ideas are encouraged."*
           ii. Brainwriting: *"Brainwriting, which is what you will be doing today, is a form of brainstorming. The difference is that in brainstorming, ideas are usually expressed aloud, whereas in brainwriting, ideas are written down. Today you will be recording ideas on notecards. The way that today's session will work is that just before you begin generating ideas; I will give you the topic.*
               1. ***Individual:*** *When I say 'begin' you will flip over the top notecard in front of you and there will be a sub-topic of the larger topic written at the top of the note card. You'll think of an idea related to the sub-topic and write that idea, legibly, on the notecard. Once you've finished writing that*

idea, you'll flip that card facedown to start a discard pile. Then you'll flip over a new notecard. That notecard will also have a sub-topic written at the top. Read the sub-topic and generate an idea related to that sub-topic. When you've finished with that idea, flip that card facedown adding it to your discard pile. Continue doing this until you've used all of the notecards in your pile. Try your <u>best</u> to generate an idea for each notecard; however, if you find that you are completely <u>unable</u> to think of another idea, put an 'X' on the card and flip the card over facedown. It is important that you record an 'X' if you are unable to generate an idea so that the researcher knows that you did not inadvertently miss that card, but that you were unable to think of idea for that card. Now let's look at an example together."

2. ***Interacting:** When I say 'begin' you will flip over the top notecard in front of you and there will be a sub-topic of the larger topic written at the top of the note card. You'll think of an idea related to the sub-topic and write that idea, legibly, on the note card in the space next to the number 1. Once you've finished writing that idea, you'll pass that card, facing up, in a <u>clockwise</u> direction to another member of your group. So, think of the hand on a clock, and the direction that they move, and take a moment now to figure out who you will pass that card to. Does anyone have any questions about which direction to pass their card? The other members of your group will have also been recording an idea so the person on your other side will be passing a card with their idea already written on it to you. When you receive that card: 1st) read the sub-topic at the top of the card, 2nd) read the idea that the other person has already written, and 3rd) think of and add your own idea related to that sub-topic in the space next to number 2. When you've finished writing that idea, pass that card, now containing two ideas, facing up, in a clockwise direction to the next member – the same member you passed to before. The next card you receive will have two ideas already written on it. When you receive that card: 1st) read the sub-topic at the top of the card, 2nd) read the 1st idea written on the card,*

3rd) read the second idea written on the card, and 4th) think of and add your own idea related to that sub-topic in the space next to number 3. That card is now considered full. Place that card, facing down, in the middle of the group so start a discard pile. All used cards will be stacked facing downward in this discard pile. After you've discarded the used card, flip over a new notecard, read the sub-topic at the top, think of and record a new idea for that notecard. This will begin the process over again. This process will continue until all notecards have been used and placed face down in the middle of the group. Try your <u>best</u> to generate an idea during each of your turns; however, if you find that you are completely <u>unable</u> to think of another idea, you may record an 'X' in the spot where you would have written the idea. It is important that you record an 'X' if you are unable to generate an idea so that the researcher knows that your turn was not inadvertently missed, but that you were unable to think of an idea for that turn. Now let's look at an example together."

   iii. Examples
- 1. Topic – "Think of ways to raise money for a local charity."
- 2. Sub-topics: *"Subtopics for this topic might include 'sales' and 'events'."*
- 3. Example ideas: *"So if your note card was labeled 'sales' you might write, 'have students donate baked goods for a bake sale' or 'have students donate books for a used book sale.' If your note card was labeled 'events' you might write, 'have students organize a 5K' or 'have local bands host a charity concert for the event.' Notice that each idea is specific. Don't just write 'bake sale,' be specific about who is involved and what they are doing – so, 'have <u>students</u> [the who] donate baked goods [the what] for a bake sale [for what purpose].' Notice that for each idea in the 'sales' subtopic, something is being sold. For each idea in the 'events' subtopic, an event is being hosted where people will come together but no goods are being sold. So make sure that your idea is <u>related to the subtopic</u>, and <u>specific about who and what is involved</u>. Finally, please make sure to write your ideas as <u>legibly</u> as*

> *possible. Print is preferable, but if your write more legibly in cursive, you may use cursive. Just make sure your ideas are readable."*

      iv. Any questions about today's task? [E: Wait a moment to give Ps a chance to think of questions.] Does everyone understand what brainwriting is and how to record ideas? [E: Wait again]

  b. Post-Session Questionnaire: *"After you (your group) are all finished with all of your notecards, just wait quietly in your seat. Everyone will finish up around the same time. Once everyone is finished, I will pass out the post session questionnaires. Please fill out the questionnaire completely and carefully. When you are finished, wait quietly in your seat. It will only take everyone a couple of moments to fill out the questionnaire. Once everyone is finished, I will collect the questionnaires, pass out a debriefing letter and dismiss you."*

C. Final Items

  a. Final Reminders: *"Remember, there is no talking during any part of today's session. If at any time you have a question or need assistance, please raise your hand and I will come to you. For instance, if your pen quits working, or you forget what to do next, do not ask another participant; just raise your hand quietly."*

  b. Topic: "*Today's topic is, 'Think of ways to improve Loyola University Chicago'.*"

  c. Begin: "*You may now begin*

APPENDIX B

POST-SESSION QUESTIONNAIRE

Post-Session Questionnaire

Please answer <u>every</u> question. Your answers are important, so please answer carefully

and honestly. Circle <u>only one number</u> for each question.

1. How clear were the instructions given at the beginning of the session?

Not at all clear   0   1   2   3   4   Very clear

2. How well did you feel that you understand what to do?

Not well at all   0   1   2   3   4   Very well

3. How enthusiastic was the experimenter?

Not at all enthusiastic 0   1   2   3   4   Very enthusiastic

4. How important do you feel it was to the experiment that you do your best on this

    task?

Not at all important   0   1   2   3   4   Very important

5. How much did you enjoy this task?

Not at all   0   1   2   3   4   Very much

6. How well do you feel that you did on this task?

Not well at all   0   1   2   3   4   Very well

7. How difficult was it to keep on generating ideas?

Not difficult at all   0   1   2   3   4   Very difficult

8. How often were you <u>unable</u> to generate ideas?

Never   0   1   2   3   4   Very often

9. How often did an idea that you previously generated occur to you again?

Never   0   1   2   3   4   Very often

10. How many additional ideas per category do you think you could have generated?

_____ (Please write in a number).

References

Barki, H., & Pinsonneault, A. (2001). Small group brainstorming and idea quality: Is electronic brainstorming the most effective approach? *Small Group Research*, *32*(2), 158-205. doi: 10.1177/104649640103200203

Briggs, R. O., & Reinig, B. A. (2007, January). In R. Sprague (Chair). *Bounded ideation theory: A new model of the relationship between idea-quantity and idea-quality during ideation*. Paper presented at 40th Hawaii international conference on system sciences, Waikoloa, Hawaii. Retrieved from http://www.hicss.hawaii.edu/hicss_40/decisionbp/01_05_05.pdf

Brown, V. R., & Paulus, P. B. (2008). Making group brainstorming more effective: Recommendations from an associative memory perspective. *Current Directions in Psychological Science*, *11*(6), 208-212.

Burton, G. E. (1987). The "clustering effect": An idea-generation phenomenon during nominal grouping. *Small Group Research*, *18*(2), 224-238. DOI: 1177/104649648701800206

Coskun, H., Paulus, P. B., Brown, V., & Sherwood, J. J. (2000). Cognitive stimulation and problem presentation in idea-generating groups. *Group Dynamics: Theory, Research, and Practice*, *4*(4), 307-329. DOI: 10.1037/1089-2699.4.4.307

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

Goldenberg, O., Wiley, J., & Larson, J. J. R. (June 01, 2013). Goal Instructions, Response Format, and Idea Generation in Groups. *Small Group Research, 44,* 3, 227-256.

Lamm, H., & Trommsdorff, G. (1973). Group versus individual performance on tasks requiringideational proficiency (brainstorming): A review. *European Journal of Social Psychology*, *3*(4), 361-387.

Larson, J. R., (2010). *In search of synergy in small group performance.* New York, NY: Psychology Pr.

Nagasundaram, M., & Dennis, A. R. (1993). When a group is not a group: The cognitive foundation of group idea generation. *Small Group Research*, *24*(4), 463-489. doi: 10.1177/1046496493244003

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review*, *10*(3), 186-213. Retrieved from http://www.grouplab.org/psy840/NijstadStroebe.2006.pdf

Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. M. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of Experimental Social Psychology*, *38*, 535-544. Retrieved from http://www.sciencedirect.com

Parnes, S., & Meadow, A. (1959). Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *Journal of Educational Psychology*, *50*(4), 171-176.

Paulus, P. B., Kohn, N. W., & Arditti, L. E. (2011). Effects of quantity and quality instructions on brainstorming. *Journal of Creative Behavior*, *45*(1), 38-46.

Paulus, P. B., & Yang, H. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes*, *82*(1), 76-87. DOI: 10.1006/obhd.2000.2888

Reinig, B. A., & Briggs, R. O. (2008). On the relationship between idea-quantity and idea-quality during ideation. *Group Decision and Negotiation*, *17*, 403-420. DOI: 10.1007/s10726-008-9105-2

Reinig, B. A., Briggs, R. O., & Nunamaker, J. F. (2007). On the measurement of ideation quality. *Journal of Management Information Systems*, *23*(4), 143-161. DOI: 10.2753/MIS0742-1222230407

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2007). Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology*, *43*, 933-946. DOI: 10.1016/j.jesp.2006.10.014

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of the measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*(1), 45-48. Retrieved from http://graphics.tx.ovid.com.flagship.luc.edu/ovftpdfs/FPDDNCIBPFCNFC00/fs046/ovft/live/gv002/00006000/00006823-197107000-00004.pdf

Stroebe, W., Nijstad, B., & Rietzschel, E. (2010). Beyond productivity loss in
 brainstorming groups: The evolution of a question. In M. Zanna & J. Olson
 (Eds.), *Advances in Experimental -Social Psychology* (pp. 157-203). DOI:
 10.1016/S0065-2601(10)43004-X

Ward, T. B. (1994). Structured imagination: The role of conceptual structure in exemplar
 generation. *Cognitive Psychology, 27*, 1-40. Retrieved from
 http://www.sciencedirect.com.flagship.luc.edu/science/journal/00100285/27/1

VITA

Amanda Egan was born and raised in Columbus, Ohio. Before attending Loyola University Chicago, she attended Cedarville University, in Cedarville, Ohio where she earned her Bachelor of Arts in Psychology, in 2005. From 2004 to 2011, she worked as an intervention specialist with students in grades kindergarten through 12th grade at Madison Christian School in Groveport, Ohio.

While at Loyola, Egan worked with Dr. James Larson in his lab studying groups and individuals working in teams. She also worked with Dr. Arthur Lurigio on a variety of projects including research on probationers with mental illness as well as violent juvenile offenders.

Currently, she is working for the Graduate School at Loyola University Chicago while she continues to pursue her Ph.D. in Applied Social Psychology at Loyola.