Oregon Undergraduate Research Journal
15.1 (2019) ISSN: 2160-617X (online)
blogs.ourj.uoregon.edu

The Role of Originality in Retrieval from Long-Term Memory: Relations Between Fluency, Originality, Working Memory Capacity, and Crystallized Intelligence

Dillon Murphy*, Psychology

ABSTRACT

To better understand the cognitive processes necessary for successful verbal fluency (a measure of retrieval from long-term memory) performance and why individuals differ in performance, the present study (n = 148) examined relations between fluency, originality, corrected fluency, working memory capacity, and crystallized intelligence. Results demonstrated that fluency significantly and positively correlated with originality; however, many of the relations between fluency, originality, corrected fluency, working memory capacity, and crystallized intelligence varied across the different category fluency tasks (animals and supermarket items). Additionally, an examination of the output position (order) of recalled items indicated that original items tended to be emitted towards the end of the recall period. Recalling common items prior to unique items may serve as a strategy by which participants begin their search through long-term memory by focusing on the most easily accessible items before emitting less accessible items. Indeed, individuals who increasingly recalled common items before unique items tended to recall more items overall, but this finding also varied across the fluency tasks. Collectively, the results suggest that originality, working memory capacity, crystallized intelligence, and output position are all factors that should be taken into consideration when accounting for variation in verbal fluency performance.

1. INTRODUCTION

Verbal fluency, the number of items a participant can generate from long-term memory (LTM) for a given category, can be an excellent measure of retrieval ability. When accessing items from LTM, many factors affect the search process, including strategy use, working memory, and task-related knowledge. For example, search strategies such as clustering items and switching between clusters have been shown to increase performance in fluency tasks (Gruenewald & Lockhead, 1980; Troyer, Moscovitch, & Winocur, 1997). Individuals may also differ in working memory

* Dillon Murphy is an undergraduate psychology student at the University of Oregon. Dillon worked as a research assistant in the Social Cognition Lab led by Dr. Sara Hodges, the Personality and Social Dynamic Lab led by Dr. Sanjay Srivastava, the Kuhl Memory and Attention Lab led by Dr. Brice Kuhl, and the Memory, Attention, and Individual Differences Lab led by Dr. Nash Unsworth where he also conducted an honors thesis. Following graduation, Dillon will be enrolling in the Cognitive Psychology Ph.D. program at the University of California Los Angeles to work with Alan Castel, Bob and Elizabeth Bjork, Barbara Knowlton, and Jesse Rissman. His research interests include attention, intelligence, short-term versus long-term memory, the capacity of our memory system, false or distorted memories, reasons for memory loss, and how these things change over the course of life among other diverse topics. Please direct correspondence to dmurphy1903@gmail.com.

capacity (WMC) as well as levels of knowledge, and these differences can affect retrieval performance (Hughes & Bryan, 2002; Rosen & Engle, 1997). That is, those with high WMC and increased task-related knowledge demonstrate better verbal fluency. The uniqueness of recalled responses can be an additional predictor of performance (e.g., Dumas & Dunbar, 2014; Silvia, 2008) and may also be related to the factors of how individuals differ in their retrieval from LTM.

1.1 FLUENCY

Research has indicated that in semantic verbal fluency tasks, participants use a two-stage cyclical process to guide their search of LTM (Gruenewald & Lockhead, 1980). For instance, in an animal fluency task, participants typically begin by searching for general categories, or clusters, of animals (e.g., pets), then begin searching for items within these clusters (e.g., cat, dog, hamster, etc.) until they have exhausted the items in said clusters. They repeat this process until all clusters have been exhausted (Gruenewald & Lockhead, 1980). Building off of this model, Troyer, Moscovitch, and Winocur (1997) referred to clustering as the generation of items within categories and referred to switching as the generation of new categories. They found both factors to be important determinants of performance. They also discovered age-related differences in overall performance and credited older adults' poorer performance to a deficit in switching but found no age differences in clustering. Unsworth, Spillers, and Brewer (2011) found a similar effect in healthy young adults. They suggested that although both clustering and switching are related to fluency, switching was more highly related to performance, indicating that the ability to self-generate category cues may be more important than the number of items emitted in each category.

1.2 WORKING MEMORY CAPACITY

Working memory, short-term memory responsible for temporarily holding information available for processing, has been found to be a predictor of high order cognitive abilities and performance in a wide variety of real-world cognitive tasks such as reasoning (Kyllonen & Christal, 1990), writing (Benton, Kraft, Glover, & Plake, 1984), reading comprehension (Daneman & Carpenter, 1980), and complex learning (Kyllonen & Stephens, 1990). When discussing working memory "capacity," researchers are referring to individual differences in storage capacity and the ability to focus attention. In terms of fluency, measures of WMC have been shown to relate to fluency scores (e.g., Rosen & Engle, 1997; Unsworth et al., 2011). Specifically, high WMC individuals typically recall more items and more clusters of items while also recalling items at a faster rate (Unsworth, Brewer, & Spillers, 2013). These differences may be due to high WMC individuals being more efficient at searching their memory than low WMC individuals, partly because high WMC individuals are better able to select and utilize appropriate search strategies to guide retrieval via self-generation of category cues and individual items within said categories (Unsworth et al., 2011).

1.3 CRYSTALLIZED INTELLIGENCE (gC)

Vocabulary tests have long been used as measures of crystallized intelligence (gC), the ability to use one's learned knowledge (Cattell, 1943). Differences in vocabulary knowledge have been found to account for fluency scores in that better fluency performance is associated with better vocabulary (e.g., Hughes & Bryan, 2002; Troyer et al., 1997; Unsworth et al., 2011). Individuals with better vocabulary may use this knowledge to enhance performance via associative links between items so that more words are generated before each category is exhausted. In terms of strategy use, research suggests that vocabulary is associated with clustering but not switching (Unsworth et al., 2011). This indicates that the associative links between words can increase output within clusters but do not assist in generating cues for new categories. Clearly, verbal fluency performance is driven both by WMC via strategic search and vocabulary via associative links between words.

1.4 ORIGINALITY

While fluency scores account for the grand total number of items that a participant generates for a given task, these scores do not account for the uniqueness of the items. Originality is the relative uniqueness of the generated exemplars. In regards to fluency, researchers have found fluency and originality to be significantly and positively related (e.g., Dumas & Dunbar, 2014; Silvia, 2008). Thus, those who recall more items overall tend to be more unique in their responses. There is some disagreement regarding the strength and direction of the relationship between fluency and originality (e.g., Hocevar, 1979); however, a significant positive relationship is typically found (e.g., Dumas & Dunbar, 2014), and these differences likely arise due to the method used to score originality.

Originality has been operationalized in many ways, such as using a panel of raters (e.g., Hocevar, 1979), participant self-report (e.g., Silvia, 2011), statistical infrequency, (e.g., Silvia, 2008), and semantic analysis (e.g., Dumas & Dunbar, 2014). Each of these methods has unique strengths and weaknesses. For example, originality scores generated by a panel of raters tend to be subjective; scores for the same word can vary significantly even when raters are highly trained (e.g., Sternberg, 2006). In self-report methods, participants may have good insight into their own minds and can therefore judge their originality. However, they are not highly trained and therefore lack an advanced understanding of the operationalization of originality, biasing this method of measurement.

As opposed to subjective measures, more objective measures (such as semantic analysis or statistical infrequency) can be more effective. In semantic analysis, the originality of ideas is scored through the use of semantic networks. Specifically, the greater the semantic distance of an exemplar from a cue, the higher the originality score it will receive. While semantically derived originality scores are not sample-dependent, semantic analysis is best suited for divergent thinking tasks (e.g., Dumas & Dunbar, 2014). When scoring originality via statistical infrequency, if a participant is the only individual in the entire sample to generate a particular item, this item would score highly for originality while more common items would receive lower scores. Although free of subjectivity, this method is highly sample dependent. The same exemplar generated by different participants in different samples could receive vastly different originality scores. Additionally, some researchers have found this method to confound originality scores with fluency scores (e.g., Hocevar, 1979). That is, one might argue that as the size of the response set increases, so does the likelihood of the set containing original items. To improve discriminant validity, corrected fluency scores calculate the total number of items recalled minus the number of original items recalled. These adjusted fluency scores help control for correlations due to heightened total output and help fluency and originality to be measured as separate constructs. Given the various strengths and weaknesses of these methods, the statistical infrequency method was used to measure originality in this study.

The goal of the current study was to gain a better understanding of the dynamics of retrieval and the relationship between fluency, originality, WMC, and gC. To examine these issues, we administered several tasks where participants performed a number of fluency, WMC, and gC measures. As suggested by prior research (e.g., Unsworth et al., 2011), we expect to replicate the finding that fluency scores correlate with WMC and gC. Using statistical infrequency as an index of originality, we should be able to track the uniqueness of participants' responses as well as the recall order of items. Specifically, the cognitive correlates of originality will be examined in order to better elucidate the processes that are needed for successful performance and why individuals differ in their performance. We suggest that originality, WMC, and gC all positively predict verbal fluency performance. We also expected fluency and originality to positively correlate and for this relationship to increase as the metric of originality is expanded. Additionally, limited research suggests a positive association between originality and gC (Silvia, 2008), but we expected to find no such relationship with WMC, as evidence suggests WMC affects the quantity of responses but not the items. Lastly, we expect original items to have a later average output position than common items. We suspect that participants will first recall common, easily accessible items and eventually transition to recalling unique, less accessible items once easily accessible items are exhausted from LTM.

2. METHOD

2.1 PARTICIPANTS

Participants were 148 undergraduate students recruited from the University of Oregon Human Subjects Pool. Participants were between the ages of 18 and 35 years and received course credit for their participation. Participants were tested individually in a laboratory session lasting approximately 2 hours.

2.2 MATERIALS

SEMANTIC FLUENCY

Participants completed two fluency tasks with different categories. In the first task, participants were instructed to retrieve as many exemplars from the category "animals" as possible in 5 minutes. The participants were informed that they could retrieve the items in any order that they wished; they were required to type in each item and then press Enter to record the response. Participants were instructed that they needed to keep trying to retrieve items throughout the entire 5-minute period. In the second task, the instructions were identical except the category was "supermarket items." The animal and supermarket tasks were selected because they have been shown to be related to other fluency measures (e.g., Unsworth et al., 2011). In fluency tasks, participants typically emit items rather quickly in the beginning, but their rate of responses slows considerably as time passes (Crowe, 1998). As a result, many researchers have used fluency tasks with varying lengths (e.g., 1 minute: Unsworth et al., 2011; 5 minutes: Unsworth et al., 2013; 15 minutes: Rosen & Engle, 1997). However, in the 1 minute version of this task, it is possible that participants do not have enough time to fully exhaust their clusters or the items within each cluster. In this study, participants were given 5 minutes to recall items as the extra 4 minutes should have given them plenty of time to exhaust each category and maximize their output.

When scoring responses for fluency in the animal task, any attempt to emit a type of living being, real or fake, was counted as correct. In the supermarket task, any item that could be located and purchased inside a typical supermarket was considered correct. Errors were categorized as repetition (emitting the same item more than once), semantic repetition (e.g., lion and lioness, soda and pop), not an exemplar (e.g., mammals, carbohydrates), and does not exist (submissions that were not animals or items found in a typical supermarket). Within the data coding process, participants' incorrect responses due to spelling errors or typos were corrected within reasonable doubt (e.g., the response "sal" followed by the response "pepper" was interpreted as "salt"). Additionally, different responses of the same exemplar were corrected to the official name of the exemplar and coded as a repetition error (e.g., "arctic fox" and "arctic snow fox" were both counted as "arctic fox" since they are semantically identical). Participants received separate fluency scores for the animal and supermarket tasks in addition to a cumulative fluency score amounting from both tasks. Task-specific fluency scores were calculated by summing the total number of correct responses for each task, while cumulative fluency scores were calculated by summing each task-specific fluency score.

ORIGINALITY

Originality scores were calculated by scoring participants based on their number of unique responses. Based on the recommendations of Milgram and Milgram (1976), for most analyses in this study, original items were considered as those that were emitted by less than 5% of the sample (the number of items per participant that were emitted 7 times or less out of the 148 included participants). Similar to fluency scores, participants received separate originality scores for each task in addition to a cumulative originality score.

CORRECTED FLUENCY

Corrected fluency scores were calculated by subtracting the participants' originality score from their fluency score. For example, if a participant had a fluency score of 50 and an originality score of 10, their corrected fluency score would be 40.

OUTPUT POSITION

In the fluency tasks, each response was tracked via output position. The difference between the mean output position for common and original items was calculated for each participant to assess at what point during the tasks original items were typically emitted. For example, if a participant's mean output position for common items was 20 and the mean output position for original items was 30, that individual's output position mean difference would be -10 (indicating that the participant recalled unique items towards the end of the task). Incorrect responses were not included in mean output position calculations. Additionally, participants with no original items in their response set were excluded from output position analysis.

WORKING MEMORY CAPACITY

Operation span. Participants solved a series of math operations while trying to remember a set of unrelated letters that were presented for 1 second each. Immediately after a letter was presented, the next operation was presented. Three trials of each list length (3-7) were presented, with the order of list lengths varying randomly. At recall, letters from the current set were recalled in the correct order by clicking on the appropriate letters. For all of the span measures, the score was the proportion of correct items in the correct position. The operation span task has been shown to have both good internal consistency and test-retest reliability (Unsworth, Redick, Heitz, Broadway, & Engle, 2009; see Unsworth, Heitz, Schrock, & Engle, 2005, for more task details).

Symmetry span. Participants were required to recall sequences of red squares within a matrix while performing a symmetry judgment task. In the symmetry judgment task, participants were shown an 8 × 8 matrix with some squares filled in black. Participants decided whether the design was symmetrical about its vertical axis. The pattern was symmetrical half of the time. Immediately after determining whether the pattern was symmetrical, participants were presented with a 4 × 4 matrix with one of the cells filled in red for 650 milliseconds. At recall, participants recalled the sequence of red-square locations in the preceding displays in the order that they had appeared by clicking on the cells of an empty matrix. There were three trials of each list length, with list length ranging from 2 to 5 (Unsworth et al., 2009).

Reading span. Participants were required to read sentences while trying to remember the same set of unrelated letters as in the operation span. For this task, participants read a sentence and determined whether or not the sentence made sense. Half of the sentences made sense, and the other half did not. Nonsense sentences were made by changing one word from an otherwise normal sentence (e.g., "During the final week of spaghetti, I felt like I was losing my mind."). After participants had indicated whether the sentence made sense or not, they were presented with a letter for 1 second. At recall, the letters from the current set were recalled in the correct order by clicking on the appropriate letters. There were three trials of each list length, with list length ranging from 3 to 7 (Unsworth et al., 2009).

For a composite WMC score, scores for the three complex span tasks were z-transformed for each participant. These z scores were then averaged together.

VOCABULARY

Synonym vocabulary. Participants were given 10 vocabulary words and were required to select the best synonym (out of five possible choices) that best matched the target vocabulary word (Hambrick, Salthouse, & Meinz, 1999). Participants were given 2 minutes to complete the 10 items. A participant's score was the total number of items solved correctly.

Antonym vocabulary. Participants were given 10 vocabulary words and were required to select the best antonym (out of five possible choices) that best matched the target vocabulary word (Hambrick et al., 1999). Participants were given 2 minutes to complete the 10 items. A participant's score was the total number of items solved correctly.

A participant's composite vocabulary (gC) score was the total number of items solved correctly. Scores were calculated by summing the scores from the two tasks.

2.3 PROCEDURE

After obtaining informed consent, participants completed operation span, symmetry span, reading span, animal fluency, supermarket fluency, synonym vocabulary, and antonym vocabulary tasks. As is typically done in individual differences studies of this type, all participants performed the tasks in the same fixed order in order to avoid the confounding of individuals with a particular task order that would complicate individual differences analyses due to an increase in error variance (Salthouse & Babcock, 1991). At the completion of the study, participants were thanked for participation, debriefed about the goals of the study, and provided contact information in case of further questions.

3. RESULTS

A total of 34 participants were excluded from analysis for having incomplete data, obvious lack of effort, or failure to follow task instructions (leaving 148 individuals for analysis). Descriptive statistics for all measures can be seen in Table 1 and the correlations between animal fluency, supermarket fluency, cumulative fluency, task-specific originality, cumulative originality, task-specific corrected fluency, cumulative corrected fluency, WMC, gC, and task-specific output position mean difference are shown in Table 2.

An examination of the results suggests a number of findings. First, fluency in the animal task correlated with fluency in the supermarket task (r = .561, p < .001). Also, originality in the animal task correlated with originality in the supermarket task (r = .513, p < .001). This demonstrates the validity of both measures and suggests that these measures of fluency may be domain-general and not limited to animals or supermarket items, but many other categories of information. As mentioned previously, originality was operationalized as items recalled by less than 5% of participants (by 7 people or fewer, out of the 148 included participants). At this level, cumulative originality significantly correlated with cumulative fluency (r = .687, p < .001) in addition to task-specific fluency (animal task: r = .685, p < .001; supermarket task: r = .667, p < .001). This demonstrates that the uniqueness of responses relates to the total number of responses and that this relationship holds true across domains.

Table 1. *Descriptive statistics for all measures.*

	Mean	Median	Std. Deviation	Minimum	Maximum
zWMC	.07	.15	.73	-2.20	1.16
Composite Vocabulary	7.54	7.00	3.46	1.00	18.00
Cumulative Fluency	104.80	105.00	20.40	53.00	170.00
Animal Fluency	44.43	45.00	10.78	20.00	75.00
Supermarket Fluency	60.34	60.00	12.30	33.00	95.00
Cumulative Corrected Fluency	84.33	84.50	14.95	38.00	117.00
Animal Corrected Fluency	38.29	39.00	8.23	19.00	58.00
Supermarket Corrected Fluency	46.04	46.50	9.18	19.00	66.00
Cumulative Originality	20.45	17.50	12.03	2.00	75.00
Animal Originality	6.14	5.00	4.94	0.00	30.00
Supermarket Originality	14.30	13.00	8.73	1.00	46.00
Animal Errors	1.71	1.00	2.47	0.00	23.00
Supermarket Errors	1.76	1.00	2.45	0.00	13.00
Total Errors	3.47	3.00	3.64	0.00	26.00
Animal Repetition Errors	1.07	1.00	2.11	0.00	22.00
Supermarket Repetition Errors	.59	0.00	.84	0.00	4.00
Cumulative Repetition Errors	1.66	1.00	2.36	0.00	23.00
Animal Semantic Errors	.64	0.00	1.22	0.00	6.00
Supermarket Semantic Errors	1.17	0.00	2.17	0.00	13.00
Cumulative Semantic Errors	1.81	1.00	2.63	0.00	14.00
Animal Common Average Output Position	22.55	22.76	5.29	10.21	36.51
Animal Original Average Output Position	30.29	30.08	8.86	11.00	51.87
Animal Output Position Mean Difference	-6.82	-7.37	7.53	-22.22	16.27
Animal Average Output Position of Errors	26.91	26.50	13.70	1.50	60.00
Animal Average Output Position of Correct Items	26.61	24.31	5.69	13.00	42.65
Supermarket Common Average Output Position	29.79	29.09	6.17	13.79	46.67
Supermarket Original Average Output Position	38.02	37.63	9.04	18.38	58.42
Supermarket Output Position Mean Difference	-8.23	-8.46	7.77	-26.00	13.06
Supermarket Average Output Position of Errors	35.35	35.30	16.20	2.00	76.14
Supermarket Average Output Position of Correct Items	32.50	31.86	6.53	17.06	49.75

Originality and its relation to fluency was evaluated across multiple levels of originality (items mentioned once, twice, three times, etc. in the entire sample). Cumulative fluency and cumulative originality strongly positively correlated, and the correlation coefficient increased as the level of originality increased (all p's $< .001$). As previously discussed, corrected fluency scores were calculated by subtracting individuals' originality scores from their fluency scores. When accounting for corrected fluency, a significant correlation with originality was still found in the animal task ($r = .297$, $p < .001$) but not in the supermarket task ($r = -.057$, $p = .491$) or cumulatively ($r = .133$, $p = .106$). This may be due to the nature of the tasks, as in the supermarket task almost a quarter (23.86%) of the total responses in the sample were considered unique as opposed to just 14.04% in the animal task. Due to the vast number of unique responses, fluency and originality in the supermarket task may lack discriminant validity.

Table 2. *Correlations for fluency, originality, WMC, vocabulary knowledge, and output position mean difference.*

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1) Cumulative Fluency	–												
2) Cumulative Originality	.687***	–											
3) Cumulative Corrected Fluency	.811***	.133	–										
4) Animal Fluency	.866***	.548***	.741***	–									
5) Animal Originality	.620***	.782***	.216**	.685***	–								
6) Animal Corrected Fluency	.763***	.249**	.841***	.899***	.297***	–							
7) Supermarket Fluency	.899***	.660***	.696***	.561***	.428***	.478***	–						
8) Supermarket Originality	.597***	.936***	.061	.368***	.513***	.175*	.667***	–					
9) Supermarket Corrected Fluency	.638***	-.006	.875***	.401***	.087	.474***	.706***	-.057	–				
10) zWMC	.247**	.213**	.166*	.239**	.146	.225**	.200*	.210*	.068	–			
11) Composite Vocabulary	.168*	.156	.104	.318***	.333***	.218**	.000	.027	-.026	.232**	–		
12) Animal Output Position Mean Difference	-.326***	-.210*	-.269**	-.259**	-.219**	-.206*	-.309***	-.166*	-.251**	-.013	.025	–	
13) Supermarket Output Position Mean Differenc	-.143	.095	-.271***	-.132	.012	-.181*	-.121	.123	-.279***	-.069	-.021	.068	–

Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$

Next, fluency and gC, as measured by vocabulary, are likely related due to certain individuals having a larger knowledge base to draw from. As mentioned above, participants typically retrieve items in clusters, and cluster size is determined in part by the knowledge that individuals bring to each task. The more prior knowledge an individual has in a given domain, the larger the cluster should be. For example, in two individuals who are the same in every category except for their gC levels , the higher gC individual would likely perform better in fluency, as they have greater access to exemplars in their LTM. Consistent with this theory, gC correlated with cumulative fluency ($r = .168$, $p = .041$) and animal fluency ($r = .318$, $p < .001$), but not supermarket fluency ($r = .000$, $p = .998$). This could be the result of the supermarket task involving more episodic memory retrieval and the animal task involving more semantic memory retrieval. For example, when searching for and selecting categories of supermarket items, participants may not rely on their semantic knowledge, but instead imagine themselves in certain aisles or recall a recent receipt at the grocery checkout. On the contrary, most individuals do not have many episodic memories for numerous different kinds of animals. Thus, participants likely use mostly semantic knowledge in the animal task while relying on additional episodic retrieval cues in the supermarket task.

While individual differences in knowledge of animals and supermarket items should be minimal, high WMC individuals typically score higher for fluency because they are able to generate more clusters of items and more items within each cluster, while the low WMC individuals are less able to access these items. Consistent with prior work (Rosen & Engle, 1997; Unsworth et al., 2011; Unsworth et al., 2013), WMC significantly and positively related to cumulative fluency (r = .247, p = .003), animal fluency (r = .239, p = .003), and supermarket fluency (r = .200, p = .015).

Based on the relationship between fluency and WMC, one might theorize an additional retrieval difference between high and low WMC individuals. In addition to the common items that most participants correctly retrieve, high WMC individuals may also be able to retrieve the less commonly known and unique exemplars that low WMC individuals are unable to access from their LTM, contributing to their increased fluency performance. Consistent with this theory, WMC correlated with cumulative originality (r = .213, p = .010) and supermarket originality (r = .210, p = .010), but not animal originality (r = .146, p = .076). Although the data somewhat supports this theory, we believe that the vast size of the supermarket response set and a lack of discriminant validity is responsible for the correlation between WMC and originality in the supermarket task. The response set, based on the total number of different items recalled, for the supermarket task (1,356) is much larger than that of the animal task (559). Similarly, the number of items considered original in the supermarket task (1,091) is much greater than that of the animal task (371). Since so many of the items in the supermarket response set were considered unique, originality and fluency may be measuring similar constructs. Consistent with this idea, the correlation between supermarket fluency and WMC (r = .200, p = .015) is almost identical to the correlation between supermarket originality and WMC (r = .210, p = .010). In contrast, in the animal task there are fewer total unique items compared to the size of the response set; hence originality scores in the animal task may still be a valid measure of uniqueness. However, the correlation between WMC and originality in the animal task was non-significant, suggesting there may be no relationship between the two constructs. One of the reasons for the relationship between fluency and WMC is that individuals differ in the number of clusters and items within each cluster generated; however, what the items are and whether or not they are unique does not seem to relate to WMC.

Next, gC and originality correlated in the animal task (r = .333, p < .001) but not in the supermarket task (r = .027, p = .745) or in cumulative originality (r = .156, p = .058). Higher levels of general knowledge are typically needed to emit unique items, thus explaining the strong correlation in the animal task. Again, this relationship difference between the tasks is likely due to the nature of the tasks. The supermarket task may require more episodic memory than the animal task, resulting in less relevance for knowledge in performance as discussed above.

Given the relationships of WMC, vocabulary, and originality with fluency, these three variables were entered into a multiple regression predicting fluency. Shown in Tables 3, 4, and 5, the resulting regression analyses suggest that originality accounted for unique variance in the total number of items retrieved in both tasks as well as cumulatively. However, WMC and vocabulary no longer accounted for significant variance in cumulative fluency once cumulative originality was taken into account. Additionally, unlike the supermarket task, originality, as well as WMC in the animal task, accounted for unique variance in the total number of items retrieved. Once again, this difference in findings between tasks is likely due to the varying size of the response set and the number of items considered unique in each task.

Table 3. *Linear regression predicting cumulative fluency scores.*

Variable	*B*	*t*	*R^2*	*F*
WMC	.097	1.543		
Composite Vocabulary	.043	.688		
Cumulative Originality	.660	10.715***	.485	45.15***

Table 4
Linear regression predicting animal fluency scores

Variable	*B*	*t*	*R^2*	*F*
WMC	.128	2.087*		
Composite Vocabulary	.076	1.178		
Animal Originality	.641	10.155***	.487	45.50***

Table 5
Linear regression predicting supermarket fluency scores

Variable	*B*	*t*	*R^2*	*F*
WMC	.071	1.036		
Composite Vocabulary	-.034	-.539		
Supermarket Originality	.653	10.320***	.449	39.19***

Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$

Finally, common and original items differed in terms of when they were emitted during the recall period. Each participant was scored for the mean output position of common and original items as well as the difference between the two. A paired samples *t*-test indicated that the mean output position of common items (22.55 in the animal task; 29.79 in the supermarket task) was significantly lower than the mean output position of unique items (30.29 in the animal task; 38.02 in the supermarket task) in both tasks (animal task: $t(137) = -12.40$, $p < .001$, $d = -1.055$; supermarket task: $t(147) = -12.90$, $p < .001$, $d = -1.060$). Thus, participants typically recalled their original responses at the end of the recall period. Additionally, animal output position mean difference was associated with fluency performance ($r = -.259$, $p = .002$), but supermarket output position mean difference was not ($r = -.121$, $p = .143$). Again, this difference across task paradigms may be attributed to differences in response set size across tasks and suggests that in both tasks, participants began retrieval with common items before recalling unique items, and in the animal task, this difference was associated with the recall of more items.

4. DISCUSSION

Overall, the results supported our hypotheses and replicated previous findings, further demonstrating that numerous factors impact participants' verbal fluency performance. Research has already indicated the importance of clustering and switching (Gruenewald & Lockhead, 1980; Troyer et al., 1997; Unsworth et al., 2011), WMC (Rosen & Engle, 1997; Unsworth et al., 2013), knowledge (Hughes & Bryan, 2002; Troyer et al., 1997; Unsworth et al., 2011), and originality (Dumas & Dunbar, 2014; Silvia, 2008). However, little research has examined the underlying factors of originality and its role in successful retrieval from LTM.

Originality was positively related to WMC in the supermarket task but not in the animal task, and this difference may be due to a lack of discriminant validity. As demonstrated by previous research, WMC is typically associated with a greater volume of responses, but this does not predict the uniqueness of said responses, only the quantity. Therefore, we think it is likely that the observed positive relation between WMC and originality in the supermarket task may be attributed to Type I error. Research using a wider variety of fluency tasks is needed to better address this possibility. On the contrary, vocabulary knowledge strongly predicted originality scores in the animal task but not in the supermarket task. This finding is likely due to higher levels of general knowledge being needed to emit unique items, but again, we believe the difference in findings may be due to the nature of the tasks, as the supermarket task may involve more episodic memory retrieval than the animal task. Further research using additional fluency tasks is needed to determine the relationship between WMC, gC, and originality.

While using the count method as an index of originality may create a discriminant validity problem with fluency, one significant retrieval difference emerged. Average output position in both tasks was found to be significantly different for original versus common items, as original items tended to be recalled towards the end of the tasks. Additionally, output position mean difference was associated with fluency performance in that participants that began retrieval with common items before recalling unique items tended to recall more items overall. However, this finding was only present in the animal task, likely due to the differences in the size of the response set as discussed above. Nevertheless, this suggests two additional aspects of the retrieval process from LTM.

First, participants likely begin retrieval with the easiest, most accessible items in LTM such as dog, cat, apples, or oranges, before recalling less-accessible items such as blue-tongued skink or Whatchamacallit candy bars. While clustering and switching were not evaluated in this study, we believe the output position tended to be later for original items because individuals start with more accessible, common items in each cluster, eventually exhausting each cluster as the task continues. High fluency individuals then begin recalling additional unique items. contributing to their higher total. Since these individuals are generating more items in each cluster, the likelihood of emitting unique items increases, thus resulting in the discriminant validity problem. Corrected fluency correlations partially support this theory, as the strength of the relationship with originality decreased in both tasks. However, more research with a wider variety of tasks is needed to determine if this finding is domain-general.

Second, as discussed above, the typical search strategy of participants is to select clusters of items and then recall items within said clusters. However, since participants recalled more total items in the supermarket task (M = 60.34, SD = 12.30) than in the animal task (M = 44.43, SD = 10.78), participants likely exhausted each category faster in the animal task. In their remaining time, participants may resort to recalling miscellaneous and unique items sooner than they would in the supermarket task, resulting in higher originality scores. A follow-up study to evaluate this theory could examine recall order as well as clustering and switching to determine if unique items are recalled towards the end of each cluster or grouped together at the end of the retrieval period in fluency tasks.

While these findings are interesting, there are a few limitations in the study. For example, the count method for scoring originality may lack discriminant validity depending on the specific fluency task. In this study, fluency and originality were only evaluated in two categorical fluency tasks. A follow-up study could evaluate fluency and originality in many more tasks such as letter fluency (e.g., words that start with the letter s) and more category fluency tasks such as recalling types of fruit or different occupations (e.g., Crowe, 1998) in addition to the animal and supermarket tasks. This would allow for greater domain-general assessment and may provide insight into some of the differences found between the supermarket and animal tasks. Additionally, using a team of raters to operationalize originality alongside the count method could further demonstrate the validity of the measure.

In terms of intelligence, this study only evaluated crystallized intelligence. Another type of intelligence, fluid intelligence, is the ability to think abstractly, identify patterns and relationships, and solve problems using logic (Cattell, 1943). Future research could investigate the relationship between fluid intelligence and fluency, originality, and WMC.

Some might argue that varying levels of category-specific knowledge for animals or supermarket items may confound the results. While someone with advanced knowledge may be able to continuously emit exemplars over the entire 5 minutes and post high fluency and originality scores, a similar study by Rosen and Engle (1997) accounted for time in their analysis, and their findings were similar even when only evaluating the first minute of the task. This, in addition to the large sample size (n = 148), suggests that participants with advanced category-specific knowledge would still show similar trends and not confound the data.

A possible alternative explanation for some of this study's findings could be that high and low WMC individuals differ in mental processing speed. This suggests that given enough time, low WMC individuals would have similar fluency and originality scores as high WMC individuals. Although not evaluated in this study, we do not believe processing speed is responsible for the results. Past studies have given participants 15 minutes for the animal task and even with triple the time, high WMC individuals still outperformed low WMC individuals in fluency (e.g., Rosen & Engle, 1997).

Additionally, levels of motivation may vary between high and low WMC individuals. For example, low WMC individuals may be less motivated to perform well, leading to lower scores on WMC, gC, and fluency tasks. Again, although not measured in this study, we do not believe that the level of effort accounted for the results. Prior studies have demonstrated that differences in WMC are unlikely to be found due to differences in motivation or effort. Heitz, Schrock, Payne, and Engle (2008) evaluated participants on the reading span task and gave participants monetary rewards based on their performance. Results demonstrated that the monetary rewards increased performance equally in high and low WMC individuals, meaning high WMC individuals still outperformed low WMC individuals. If participants had overall differences in motivation, differences in all tasks would be expected. Although motivation and effort clearly impact performance, these variables typically do not affect WMC.

Lastly, this study's findings may be limited in that the directions given to participants for each fluency task were vague. As is typically done in fluency tasks (e.g., Unsworth et al., 2013), participants were simply told to come up with as many items as possible. The task did not instruct participants to be creative or original; thus, participants may not have been trying to be creative or original in their responses. A future study could conduct a task with similar conditions but instruct the participants to try to be unique and original.

In sum, while more research is needed to fully understand the constructs, the present study replicates previous findings, and the results suggest that fluency and originality, scored via statistical infrequency, are best conceptualized as distinct but positively related constructs. The order of recalled common and unique items may be an additional strategic difference in that participants typically begin their search with easily accessible items before emitting less accessible items.

ACKNOWLEDGMENTS

I would like to thank Ashley L. Miller, Matt Robison, and Nash Unsworth for input on the direction of the project, detailed comments on drafts of the manuscript, and for constructive feedback. I would also like to thank Drew Murphy and Len Murphy for their helpful contributions in coding the data. Additionally, I thank the participants of the study for their time, interest, and cooperation.

REFERENCES

Benton, S.L., Kraft, R. G., Glover, J. A., & Plake, B. S. (1984). Cognitive capacity differences among writers. *Journal of Educational Psychology, 76,* 820-834.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.

Crowe, S.F. (1998). Decrease in Performance on the Verbal Fluency Test as a Function of Time: Evaluation in a Young Healthy Sample. *Journal of Clinical and Experimental Neuropsychology, 20*, 391-401.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466.

Dumas, D. & Dunbar, K. (2014). Understanding Fluency and Originality: A latent variable perspective. *Thinking Skills and Creativity*, *14*, 56–67.

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 225–240.

Hambrick, D. Z., Salthouse, T. A., & Meinz, E. J. (1999). Predictors of crossword puzzle proficiency and moderators of age–cognition relations. *Journal of Experimental Psychology: General*, *128,* 131–164.

Heitz, R.P., Schrock, J.C., Payne, T.W., & Engle, R.W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, *45*, 119–129.

Hocevar, D. (1979). A Comparison of Statistical Infrequency and Subjective Judgment as Criteria in the Measurement of Originality. *Journal of Personality Assessment, 43,* 297-99.

Hughes, D. L., & Bryan, J. (2002). Adult age differences in strategy use during verbal fluency performance. *Journal of Clinical and Experimental Neuropsychology*, *24*, 642–654.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence, 14,* 389-433.

Kyllonen, P.C., & Stephens, D.L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences, 2,* 129-160.

Milgram, R.M., & Milgram, N.A. (1976). Creative thinking and creative performance in Israeli students. *Journal of Educational Psychology*, *3*, 255-259.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*, 211–227.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776.

Silvia, P. J. (2008). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creativity Research Journal*, *20*, 34–39.

Silvia, P. J. (2011). Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances, and consequences tasks. *Thinking Skills and Creativity*, *6*, 24–30.

Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal*, *18*, 87–98.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*, 138–146.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: the role of controlled search. *Memory & Cognition*, *41*, 242–254.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.

Unsworth, N., Redick, T.S., Heitz, R.P., Broadway, J., & Engle, R.W. (2009). Complex working memory span tasks and higher-order cognition: A latent variable analysis of the relationship between processing and storage. *Memory, 17,* 635-654.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of switching, clustering, and overall performance. *Quarterly Journal of Experimental Psychology*, *64*, 447–466.