

Investigating the Relations Among Different Measures of False Memory

Francesca Falzarano and Karen L. Siedlecki

Fordham University, Department of Psychology, Bronx, New York

ABSTRACT

Researchers have used the term false memory to describe various memory errors, including the incorporation of erroneous information into a memory, misremembering a word presented as a picture, and the construction of a detailed memory of an event that did not occur. Whether such diverse manifestations of false memory are assessing the same construct has not been evaluated. The purpose of this study is to examine the relations among a set of variables that have been used in the literature to measure false memory. The sample consisted of 112 college students who completed four false memory measures, including the commonly used Deese-Roediger-McDermott (DRM) and the misinformation tasks. Zero-order correlations indicated that there are little to no associations between false memories in the DRM and the misinformation tasks, as well as the other false memory tasks. A confirmatory factor analysis of the DRM and misinformation variables further suggested that the false memory variables share little variance in common and may not be represented by a unitary factor. Thus, the results of the current study suggest that tasks intended to measure false memory may be measuring different types of memory errors.

KEYWORDS

false memory
DRM
misinformation
convergent validity

INTRODUCTION

Most memories are distorted to some degree. When recalling an event from the past, memories can be susceptible to distortion and may involve both fictional and factual details. Schacter, Guerin, and St. Jacques (2011) and Butler, McDaniel, Dornburg, Price, and Roediger (2004) explain that human memory is not an exact replica of prior experience; instead, memory is an adaptive and malleable process that is vulnerable to various inaccuracies. False memories may reveal themselves in many ways. For example, a detail of an event may be inadvertently recalled, or the recalled event may have never been experienced at all. Roediger and McDermott (2000) describe these situations as the “fundamental errors of remembering” (p. 123). According to Leding (2012), the topic of false memory has garnered substantial

attention by researchers, the media, and laypeople alike, in part due to its prominence in childhood sexual abuse and recovered memories in psychotherapy. Despite the increase in research focusing on the topic of false memory, researchers have not arrived at a consensus regarding a specific definition of what a false memory comprises. This lack of consensus may have implications when attempting to describe and understand assessments of false memory in the laboratory. Thus, the purpose of this paper is to use a confirmatory factor analysis (CFA) approach to investigate the relationships between four different meas-

Corresponding author: Francesca Falzarano, Fordham University, Psychology Department, 441 E. Fordham Road, Bronx, NY 10458.
E-mail: ffalzarano@fordham.edu

ures that have been used in the literature to assess false memory. These measures were chosen based on Pezdek and Lam's (2007) framework, which outlines various categories of false memory paradigms that have been used by previous researchers.

Multiple Measures of False Memory

The use of several false memory assessments imply that they measure the same underlying construct. Pezdek and Lam (2007) and DePrince, Allard, Oh, and Freyd (2004) argue that false memories should only refer to entirely novel events that have never been experienced (rather than, e.g., errors in recalling details of an event, which they argue, should be categorized as a flawed memory).

Pezdek and Lam (2007) outlined six categories of laboratory paradigms that have been used in previous literature to assess false memory. These categories include: (a) whole new event planted, (b) new or changed details planted (as demonstrated in Loftus, Miller, & Burns' [1978] misinformation paradigm), (c) the Deese-Roediger-McDermott (DRM) paradigm, (d) general recognition memory, (e) source monitoring, and (f) other. Although numerous studies have used the paradigms listed in the above categories, using the term *false memory* to universally describe each laboratory phenomenon has not been without debate.

Research examining the relationships among false memory measures have either found very weak or nonexistent correlations (Bernstein, Scoboria, Desjarlais, & Soucie, 2018; Monds, Patterson, & Kemp, 2016; Nichols & Loftus, 2019; Ost et al., 2013; Patihis, Frenda, & Loftus, 2018; Zhu, Chen, Lotus, Lin, & Dong, 2013, for an overview of relevant literature, see Bernstein et al., 2018). In fact, Bernstein et al. (2018) refer to the unitary term of false memory as "more of a linguistic convenience than a useful theoretical construct" (p. 173). In the DRM, participants are presented with word lists that are semantically related to an unrepresented critical lure. When given a memory task, participants tend to erroneously remember the lure that was not presented. Alternatively, the misinformation task, which is comprised of three phases, assesses the tendency of endorsing misleading post-event information as being part of an original event (Ost et al., 2013; Takarangi, Parker, & Garry, 2006). Ost et al. (2013) propose that the DRM and misinformation tasks could be related at the source monitoring level. However, when comparing performance on the DRM and the misinformation tasks, Zhu et al. (2013) and Nichols and Loftus (2019) explain that the DRM is fundamentally different from the misinformation paradigm in terms of methodology. False memories in the misinformation paradigm are induced through external suggestion while the DRM paradigm induces semantically generated false memories by creating a situation whereby the critical lure is associated with feelings of familiarity, consequently causing false memories to be elicited. Therefore, if the misinformation effect is a consequence of being explicitly misled, then DRM-related errors should share little or no variance with misinformation-associated errors.

Similarly, Mazzoni (2002) differentiates between two types of memory distortions: naturally occurring and suggestion dependent. Naturally occurring distortions result from the malleability of the

memory system itself. An example of a naturally occurring distortion was used by Roberts (2002), who used a picture/word recall task "designed to tap a particular kind of source monitoring error thought likely to be important in a clinical setting" (p. 423). Moreover, the Childhood Experiences Questionnaire (CEQ; Lindsay, Wade, Hunter, & Read, 2004), which falls under the *other* category of false memory paradigms proposed by Pezdek and Lam (2007), examines whether individuals are able to remember highly unlikely or impossible childhood events. Lindsay et al. (2004) note that the emergence of false memories of an unlikely event can be explained by naturally occurring distortions that may stem from interference from exposure to similar events, leading to *psuedomemories*. This idea is consistent with the fuzzy trace theory, which asserts that memory consists of dual independent systems (verbatim and gist representations), and these two systems are encoded in parallel but are stored separately (Brainerd & Reyna, 2002; Reyna & Lloyd, 1997). Within this framework, false memories are believed to emerge due to the incorrect use of gist representations to recall verbatim details of an event.

Literature that has examined suggestion-dependent false memories and spontaneous false memories has yielded mixed findings. Gallo (2010) and Platt, Lacey, Iobst, and Finkelman (1998) found that distorted autobiographical memories are related to errors on the DRM paradigm. In contrast, Wilkinson and Hyman (1998) and Otgaar and Candel (2011) found no relation between spontaneous and suggestion-based false memories. One potential explanation for this could be related to how the false memories are activated. With autobiographical memory tasks, schemas may contribute to the activation of false information, while in the DRM task, the related words semantically activate a critical lure. However, Otgaar, Verschuere, Meijer, and Van Oorouw (2012) did find a significant relationship between implanted false memories and DRM false memories, although caution should be used in interpreting these findings due to the study's small sample size. Despite the fact that research has shown that the various measures discussed above may not be related, the tasks are often inclusively categorized as measures of false memory. Thus, in order to refer to false memory as a coherent construct, the relations among several measures used previously in the literature to assess false memory should be examined. Prior empirical evidence examining the relationships between false memory measures has been limited to correlational analyses. Thus, the primary goal of the current study is to extend prior work by employing a factor-analytic approach to examine whether diverse false memory paradigms demonstrate convergent validity.

The Current Study

As noted above, an array of paradigms has been used to assess false memory in laboratory conditions. The overarching goal of this paper is to examine the convergent validity of false memory measures as a unitary construct. As explained by Siedlecki et al. (2009), convergent validity is one step in evaluating construct validity. Convergent validity is assessed by determining whether the variables postulated to represent a latent construct have a significant amount of common variance, as examined through CFA. The magnitude of each loading should be

fairly large and significantly different from 0. To that end, in the current study, participants completed four false memory tasks (the DRM task, the misinformation task, a picture/word recall task, and the CEQ) which correspond with Pezdek and Lam's (2007) categories of DRM, new or changed details planted, source monitoring, and other, respectively. However, although Pezdek and Lam categorized the picture/word recall test as being part of the general recognition category, it may be better to categorize this particular false memory task as a measure of source monitoring because the task measures false memories related to the misidentification of reporting a picture as a word, and vice versa, in the form of a recall test. Thus, the current study categorized the picture/word recall task as a measure of source monitoring. Separate structural equation models were examined to evaluate the best representation of the false memory variables. If the different tasks represent a similar dimension of individual differences, then it is expected that the one-factor model will fit best. If each task represents a separate dimension of individual differences, then a multi-factor model would be a better representation of the data. Results of this study will provide additional insight into whether these various paradigms all measure the same construct.

METHOD

Participants

The sample consisted of 112 participants between the ages of 18 and 47 ($M = 21.35$, $SD = 4.53$; 79 females, 31 males, 2 not reported). The majority of participants (55%) reported their race as White, the remainder of participants reported their race as African American (4.6%), Asian (18.3%), Hispanic/Latino (10.1%), 11% reported to be more than one ethnicity, and .9% reported their race as "other". The majority of participants (60.9%) reported some college (but no degree) as highest level of education completed. Inclusion criteria included speaking English as a primary language.

Measures

THE DEESE-ROEDIGER-MCDERMOTT PARADIGM

In this task, participants were presented with 8 lists of 15 words on a computer screen. The words presented within each list were semantically related. For example, the words *nose*, *breathe*, *sniff*, and *aroma* were all semantically related to the critical lure, *smell*, which was not presented. Word lists were selected from those provided by Stadler, Roediger, and McDermott (1999) and were selected because they were associated with higher rates of eliciting a false memory effect. Participants were given instructions to pay close attention to the words and were informed that their memory for the words will be tested. Following the administration of each word list, participants completed a recall test by writing down all the words they could remember in an answer booklet. Recall of the critical lure was recorded as a false memory. False recall scores ranged between 0 and 8, and mean recall of critical lures was then calculated. Higher values indicated more critical lures falsely recalled.

THE DEESE-ROEDIGER-MCDERMOTT RECOGNITION TASK

A recognition task (modified from Salthouse & Siedlecki, 2007) which contained 24 old words, 8 critical lures, and 32 new words, was administered after all DRM lists. This task was implemented after data collection began; therefore, a total of 93 participants completed this task. The old items were words presented in the 4th, 8th, and 12th positions in the presentation, the absent critical lure from each of the 8 stimulus lists, and 3 items and 1 critical lure from the 8 lists that were not presented in the study phase. Participants indicated whether the word presented was new or old, and then were asked to report how confident they were of that report on a scale of 1 (*not confident at all*) to 5 (*very confident*). False recognition of critical lures was calculated by summing the eight original critical lures that were erroneously recognized as old. Higher values indicated higher rates of false recognition. Corrected recognition (hit rate - false alarm rate) was also calculated to assess veridical memory accuracy. Corrected recognition rates are single values and hence, cannot be examined for reliability, so hit rate was examined for reliability instead (Cronbach's $\alpha = .68$). Furthermore, to examine recognition memory, number of old words, new words, and old critical lures that participants recognized were summed and then converted into proportions.

MISINFORMATION PARADIGM

Modified from Takarangi et al.'s (2006) version of the misinformation paradigm, this task examined the likelihood that participants would attribute misleading post-event information to original event information. Participants completed three different stages of this task. In the first stage (original event), participants viewed a video on a computer screen. The video showed an electrician entering a client's home to work and while completing his work, snooping around various parts of the house and stealing items from the home. In Stage 2 (the narrative), participants were then asked to read a passage similar to what was viewed on the video. Four critical items were manipulated in the narrative in order to introduce misleading post event information. For example, in the video, the electrician went into a refrigerator and drank a Coca Cola. However, the electrician in the narrative was described as drinking a Pepsi. After a time-filled delay, in which participants completed the Creative Imagination Scale (See Appendix A), Stage 3 of the task was administered, which was a 20 question, two-alternative forced choice recognition test (as employed by Takarangi et al., 2006) regarding the details of the video that was watched. Following each forced choice question, participants were asked to provide confidence ratings on the accuracy of their responses (1 - *not confident at all* to 5 - *very confident*). Critical items from the narrative mistakenly recognized as correct were each given a score of 0 (if the response was correct) or 1 (if the response was incorrect). The total misinformation score was then calculated by summing the four items. Total scores ranged from 0 to 4, with higher numbers indicating more false memories. Veridical memory performance on the misinformation task was also examined, with larger numbers indicating a greater number of noncritical questions answered correctly.

PICTURE/WORD RECALL TASK

This task was adapted from Roberts (2002). Participants were presented with 25 pictures and 25 words on a computer screen for 3 s each, with a 1 s interstimulus interval. Prior to administration of this task, participants were instructed to keep a mental tally of the number of animate and inanimate objects that were presented. Participants were unaware their memory would be tested after the presentation of pictures and words. After the stimuli presentation, participants recorded their tally of animate and inanimate objects on the response sheet. After a 20-30 minute filled delay (in which several questionnaires were administered), participants completed a surprise memory task where they were instructed to recall all of the pictures that were previously presented in Part 1. Participants were explicitly instructed to only provide the names of the pictures presented, not the words presented. Participants were then asked to recall all of the words presented during the same task. Words falsely recalled as pictures, with possible scores ranging from 0 to 25, were classified as false memories. Conversely, pictures falsely recalled as words, with possible scores ranging from 0 to 25, were also classified as false memories. Participants' scores for both falsely recalled pictures and words ranged between 0 and 2, with higher values indicating more false memories recalled. Number of pictures and words correctly recalled were also calculated to assess veridical memory. Participants' scores ranged between 1 and 11 for pictures correctly recalled, and 0 and 12 for words correctly recalled.

CHILDHOOD EXPERIENCES QUESTIONNAIRE

The CEQ was adapted from Lindsay et al. (2004) and comprised 32 statements regarding childhood experiences (prior to the age of 12). Twenty-eight of these childhood experiences are typical events that most individuals in the U.S. are believed to have experienced at some point during their childhood (e.g., "ride on a merry go round," "have chicken pox"). Four critical items randomly inserted in the measure were considered impossible or highly unlikely events. These items included "seeing cigarette advertisements on television" (CEQ critical lure 1), which is unlikely because these were banned from the U.S. in 1971. This item was included in the original CEQ (Lindsay et al., 2004). The following critical lures were added to the CEQ for the current study: "visiting Bugs Bunny at Disney Land" (CEQ critical lure 2; adapted from Braun, Ellis, & Loftus, 2002), impossible because Bugs Bunny is not a Disney character; "having one's diaper changed prior to the age of 1" (CEQ critical lure 3) as well as "seeing a colored mobile hanging over one's crib the day after birth" (CEQ critical lure 4), both not plausible because of the infantile amnesia phenomenon. Participants rated each item on the following 5-point scale: *event never experienced*, *maybe experienced but no specific memories of it occurring*, *event was experienced but no specific memories of it occurring*, *event was experienced with weak memories*, and *event was experienced with clear memories of the occurrence*. Responses were dichotomized so that the first three scale choices were classified as no memory and the last two choices were categorized as false memories. Cronbach's α of the CEQ, excluding the four critical items, was .72.

Assessing Participant Awareness

MOOD ASSESSMENTS

The influence of demand characteristics is a notable source of discussion in the false memory literature. Thus, to minimize potential demand characteristics, the current study utilized a *red herring* technique in which participants were provided with a false but plausible rationale of what the study was about (Laney et al., 2008). In the current study, participants were led to believe that the purpose of the study was to assess the relationships between memory and mood/mood changes. This was accomplished by administering a one-item mood assessment derived from Mayer and Gaschke's (1988) Brief Mood Introspection Scale throughout the study on five separate occasions. Participants were asked to rate their current mood ranging from -5 (*very unpleasant*) to 5 (*very pleasant*).

PARTICIPANT AWARENESS

At the conclusion of the study, participants completed an open-ended question in which they were asked to indicate what they thought the study was about. Participants then completed a questionnaire in which they were provided with five alternatives to choose from and were asked to choose the option(s) that they believed best described the purpose of the study. The choices that were provided were: relationships among different types of memory and personality characteristics (which is true in many ways, and is what participants were told the study was about), the relationship between mood/mood changes and memory performance (which was the red herring previously mentioned), how creativity relates to memory errors and influences of childhood experiences on memory performance, whether different measures of false memory and memory errors are related to one another (i.e., the true purpose of the study), and "other".

Procedure

Participants were recruited as part of a larger study and were asked to complete several questionnaires, along with the four measures of false memory (mentioned above) in one visit to the laboratory that lasted approximately 90 minutes (See Appendix A for procedure order and detailed task descriptions). Because the focus of this study was to examine the relationships among different false memory paradigms, the tasks described above are limited to the measures relevant to the current analyses. Of note, this study adopted an individual differences approach. Thus, all participants received the study measures in the same order, which is a common practice in individual differences research. Specifically, Salthouse and Siedlecki (2007) note that although measures of false memory often involve the administration of different types of stimuli for each task in prior research, administering each item to all participants in the sample allows for the investigation of whether individual differences are associated with the tendency to make false memory errors.

Structural Equation Modeling

The CFAs were conducted with Amos 24.0 (Arbuckle, 2016). Full information maximum likelihood estimation was used. Several fit statistics were evaluated to assess model fit, including chi-square, comparative fit index (CFI), and root mean square error of approximation (RMSEA). While modest, our sample size ($N = 112$) is likely adequate for the simple CFA models we examine below (e.g., Wolf, Harrington, Clark, & Miller, 2013). A p value of .05 was used for all analyses.

RESULTS

The means, SD s, and percentages of the false memory and veridical memory variables are reported in Table 1. As seen in Table 1, the DRM corrected recognition variable was substantially kurtotic and was transformed using the reflect/square root method ($M = .49$, $SD = .28$, skewness = $-.21$, kurtosis = $-.49$). The transformed variable was used in all subsequent analyses.

Deese-Roediger-McDermott Performance

FREE RECALL

A paired-samples t -test was conducted to compare differences in the proportion of critical lures falsely recalled and items correctly recalled in the DRM task. Results indicated that there was a significant difference, $t(109) = -5.65$, $p < .001$, $d = .82$, such that the proportion of recall was significantly higher for items that were presented in the list ($M = .57$, $SD = .11$) compared to critical lures that were not presented ($M = .41$, $SD = .25$).

RECOGNITION

The number of old words, lures, and new words presented in the recognition task were summed and then converted into proportions. A one-way within-subjects analysis of variance (ANOVA) with the Greenhouse-Geisser correction revealed that there was a significant main effect of word type on recognition memory, $F(1.50, 137.80) = 467.86$, $p < .001$, partial $\eta^2 = .84$. Bonferroni post-hoc analyses showed that the proportion of old words recognized was significantly higher ($M = .83$, $SD = .13$) than both the proportion of new words recognized ($M = .05$, $SD = .10$; $p < .001$) and critical lures recognized ($M = .73$, $SD = .27$; $p = .01$). Additionally, the proportion of critical lures recognized was significantly higher than the proportion of new words recognized ($p < .001$), demonstrating a false memory effect.

Misinformation Performance

ACCURACY AMONG CRITICAL AND CONTROL ITEMS

Participants were assessed on their level of accuracy in recognizing the four critical items compared to the four control items. Results of a paired-samples t -test revealed that participants were significantly more accurate at recognizing control items ($M = 3.52$, $SD = .66$) than misleading critical items ($M = 2.60$, $SD = 1.13$), $t(109) = 8.15$, $p < .001$, $d = .99$.

DIFFERENCES IN CONFIDENCE RATINGS

To examine differences in participants' confidence levels for misleading and control items, a paired-samples t -test was conducted. Results indicate that participants were significantly more confident in their responses to the misleading critical items ($M = 4.23$, $SD = .60$) as compared to the control items ($M = 3.80$, $SD = .06$), $t(109) = -7.10$, $p < .001$, $d = 1.01$.

TABLE 1.
Descriptive Statistics for Memory Variables

	M	SD	Skewness	Kurtosis	Percentage of false/correct memory
False memory variables					
Falsely recalled DRM critical lures	3.30	2.02	0.15	-0.75	41.25
Proportion of falsely recalled DRM critical lures	0.41	0.25	0.15	-0.75	5.12
Falsely recognized critical lures in the DRM recognition task	5.88	2.20	-1.09	0.30	7.35
Proportion of falsely recognized critical lures in the DRM recognition task	0.73	0.27	-1.09	0.31	9.13
Critical items recalled in the misinformation task	2.60	1.13	-0.47	-0.67	64.77
Number of words falsely recalled	0.21	0.50	1.88	1.85	0.75
Number of pictures falsely recalled	0.36	0.64	1.61	1.31	1.42
CEQ unlikely/impossible events	0.31	0.55	1.60	1.65	10.33
Memory variables					
Number of DRM items correctly recalled	67.58	12.95	0.000	-0.35	56.32
Proportion of DRM items correctly recalled	0.57	0.11	0.000	-0.35	0.47
DRM corrected recognition*	0.78	0.17	-2.73	11.90	
Recognition of control items in the misinformation task	3.52	0.66	-1.82	1.62	88.00
Recognition of non-critical items in the misinformation task	14.81	1.25	-1.17	1.35	92.56
Number of words correctly recalled	4.27	1.80	1.14	1.31	11.64
Number of pictures correctly recalled	2.91	2.63	0.90	1.09	17.10
CEQ likely events	3.12	0.46	-0.13	-0.49	11.14

Note. 93 participants completed the DRM recognition task. * = untransformed variable

Picture/Word Recall Task

There was a superiority effect such that individuals were significantly more likely to correctly recall words ($M = 4.27$, $SD = 1.80$) as compared to pictures ($M = 2.91$, $SD = 2.63$), $t(109) = 4.60$, $p < .001$, $d = .60$. In terms of false memory, there was a likely floor effect, with both the mean number of words incorrectly recalled as pictures ($M = .21$, $SD = .50$) and the mean number of pictures incorrectly recalled as words ($M = .36$, $SD = .64$) being close to zero. There were no significant differences in the number of words incorrectly recalled as pictures and the number of pictures incorrectly recalled as words, $t(109) = -1.94$, $p = .06$, $d = .26$. Because of the likely floor effects, and the low reliability of the task overall, this variable was excluded from the CFAs.

Childhood Experiences Questionnaire

The impossible/highly unlikely events were dichotomized to reflect whether an item was recalled (indicating a false memory) or not recalled (indicating no false memory). For the critical item (CEQ critical lure 1) "seeing cigarette advertisements on television," 77.7% ($n = 87$) of the participants had no memory of the false event and 22.3% ($n = 25$) of the participants were categorized as having a false memory. Of the participants who have been to Disneyland in the past, 94% ($n = 78$) had no memory for the critical item "seeing Bugs Bunny at Disneyland" (CEQ critical lure 2), while 6% ($n = 5$) were categorized as having a false memory. For the item, "seeing a colored mobile hanging over one's crib the day after birth" (CEQ critical lure 3), 95.5% ($n = 107$) of the participants had no memory of the impossible event, and 4.5% ($n = 5$) of the participant responses were categorized as a false memory. For the final item, "having one's diaper changed prior to the age of 1" (CEQ critical lure 4), no participants reported a false memory for this event, so this item was not included in subsequent analyses. A mean CEQ variable was then calculated to form a composite variable. However, because this measure suffers from a potential floor effect due to the items producing very low false memory rates, along with the measure's poor psychometric properties, the CEQ was not included in the CFAs.

Relationships Among the False Memory Variables

The correlation matrix for the false memory variables is presented in Table 2. Inspection of the correlation matrix shows that only a few of the correlations were significant. For example, falsely recalled DRM critical lures were associated with number of words incorrectly recalled as pictures in the picture/word recall task ($r = .26$, $p = .006$). Additionally, falsely recalled DRM critical lures were significantly related to falsely recognized critical lures in the DRM recognition task ($r = .42$, $p < .001$).

Confirmatory Factor Analysis

The false memory variables from the picture/word task and the CEQ tasks were excluded from the final CFAs because of their poor psychometric properties (described above). The CFAs were conducted to examine the best representation of the data using variables from the DRM and misinformation tasks. However, these analyses are exploratory since our original analytic plan included examining the convergent validity of all four tasks in a series of models. Because we excluded both the picture/word and CEQ tasks, our exploratory analyses included only a one-factor and a two-factor model. First, a two-factor model was examined in which the two factors corresponded to the DRM and misinformation tasks, respectively. The DRM construct comprised the two DRM indicators (proportion of critical lures falsely recalled and proportion of critical lures falsely recognized) and the misinformation construct comprised the four individual misinformation variables. The variances of the latent factors were set to 1.0 to allow model identification. Inspection of the estimates of this model showed a negative error variance associated with proportion of critical lures falsely recalled in the DRM. To deal with this issue, the error variance for this variable was set to zero. In addition, the standardized loading of one misinformation item (Critical item 4) was essentially zero ($-.01$), so this variable was removed from the model. After these changes, the two-factor model fit well, $\chi^2 = 4.90$, $df = 5$ (ns), $\chi^2/df = .98$; CFI = .980; RMSEA = .00, AIC = 34.90. The standardized coefficients were all significantly different from zero (ranging in .42 to 1.0; see Figure 1). The correlation among the two constructs was $-.23$ ($p = .06$).

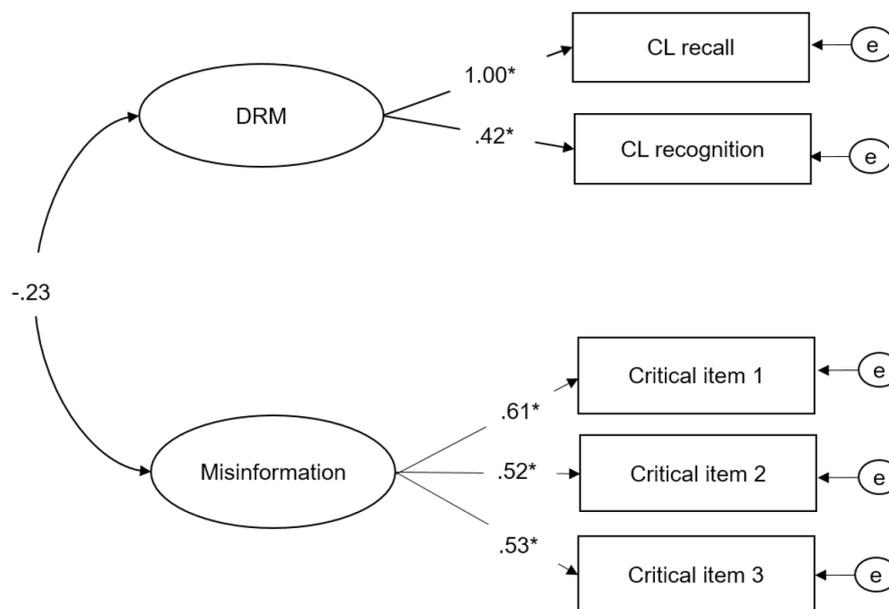
TABLE 2.

Relationships Among Measures of False Memory and Veridical Memory Performance

	1	2	3	4	5	6	7	8	9	10	11
1. Falsely recalled DRM critical lures	(.61)										
2. Falsely recognized critical lures in the DRM recognition task	.42**	(.79)									
3. Critical items recognized in the misinformation task	-.11	.02	(.42)								
4. Words incorrectly recalled as pictures	.26**	.17	-.09	(.21a)							
5. Pictures incorrectly recalled as words	.11	-.18	.08	.05	(.21a)						
6. CEQ mean critical lures	-.03	.16	.06	-.10	-.08	(.23)					
7. Number of DRM correctly recalled	-.13	-.39**	.34**	-.12	.05	-.17	(.84)				
8. DRM Corrected recognition	-.33**	-.83**	-.10	-.16	.08	-.01	.22*	(n/ab)			
9. Recognition of filler items in misinformation task	-.11	-.22*	.02	.08	-.00	-.14	.24*	.10	(.35)		
10. Number of words correctly recalled in picture/word recall task	-.03	-.04	.09	-.05	.00	-.01	.28**	.01	.18	(.21a)	
11. Number of pictures correctly recalled in picture/word task	.11	-.19	.27**	-.06	.28**	-.05	.34**	.13	.15	.05	(.21a)

Note. Numbers in parentheses are Cronbach's α values. DRM = Deese-Roediger-McDermott paradigm; CEQ = Childhood Experiences Questionnaire. a = reliability for all 50 items of the Picture/Word Recall Task. n/ab = Corrected recognition rates are single values and cannot be examined for reliability. participants completed the DRM recognition task. * = untransformed variable.

* $p < .05$, ** $p < .01$

**FIGURE 1.**

Two-factor model of the false memory variables. DRM = Deese-Roediger-McDermott; CL= critical lure; e = error term.

* $p < .05$

Next, the same five indicators loaded onto a single construct. This model fit substantially worse, $\chi^2 = 29.42$, $df = 6$ ($p < .001$), $\chi^2/df = 4.90$; CFI = .401; RMSEA = .188, AIC = 57.42. Only the DRM variables loaded significantly onto the unitary construct. The standardized coefficients for the misinformation variables were not significant and ranged from $-.02$ to $.18$.

Relationships Among False memory and General Memory Performance

Correlations assessing false memory performance and variables assessing veridical memory performance are presented in Table 2 (correlations using the untransformed DRM-corrected recognition variable can be found in Table 1 in the Supplementary Materials). Additionally, the number of noncritical items correctly recalled in the DRM paradigm was negatively correlated with falsely recognized critical lures in the DRM recognition task ($r = -.39$, $p < .001$) and the number of critical items recognized in the misinformation task was positively related to number of pictures correctly recalled in the picture/word recall task ($r = .27$, $p = .004$).

Participant Awareness

Participants answered an open-ended question in which they were asked to report what they believed was the true purpose of the study. Two coders evaluated each response, which were classified into three (non-mutually exclusive) categories: false memory, mood and memory (i.e., the red herring), and other. Overall, 9 (8%) responses were categorized as pertaining to false memory, 24 (21.4%) were categorized as pertaining to mood and memory, and 77 (68.8%) were categorized as "other".

Participants were then asked to choose which alternative (described in the Methods section) they believed described the purpose of the study. Overall, 56 participants (50.45%) selected the false memory option and 55 participants (49.55%) selected one of the four other alternatives. Independent-samples t-tests showed that those who were possibly aware of the purpose of the study did not perform differently on the false memory measures as compared to those who did not select the false memory option, with one exception. There was an effect of possible awareness on words incorrectly recalled as pictures, $t(108) = 2.55$, $p = .01$, $d = .48$, indicating that those who believed that the study may be about false memory falsely recalled more words as pictures ($M = .32$, $SD = .61$) compared to the unaware group ($M = .09$, $SD = .30$).

DISCUSSION

The purpose of the current study was to use a factor analytic approach to examine the relations among a set of variables hypothesized to reflect the same underlying construct. Some research (e.g., Lövdén, 2003) has found that false memory variables share substantial variance in common and exhibit some degree of convergent validity. However, other work has found little to no such relationship (e.g., Monds et al., 2016; Nichols & Loftus, 2019; Ost et al., 2013; Salthouse & Siedlecki, 2007; Wilkinson & Hyman, 1998; Zhu et al., 2013). For example, Nichols and Loftus (2019) examined the relationships between the DRM and misinformation tasks, along with an imagination inflation activity and found largely small to null relationships among the measures. Additionally, Salthouse and Siedlecki (2007) examined the relations among different versions of the DRM and found little relation between the false memory variables (r s ranging from $-.05$ to $.40$).

The results of the current study suggest that the false memory variables share little variance in common and are likely assessing different types of memory errors. However, within the context of latent variable analysis (which reduces the influences of unreliability), the correlation ($-.23, p = .06$) among the constructs representing the DRM and misinformation tasks was slightly larger in magnitude than the zero-order correlation among the observed variables. Thus, our results show that falsely remembering critical lures in the DRM task is negatively associated with falsely remembering information in the misinformation task. However, the relationship between the two constructs in the current study did not reach significance ($p = .06$) and thus this null result should be interpreted with caution. It is worth noting, though, that others have reported small, negative associations between the two tasks. For example, Ost et al. (2013) reported a mean correlation of $-.01$, and Calvillo and Parong (2016) reported a relationship of $-.07$ between critical lures in the DRM and misinformation tasks. A possible explanation for this negative relationship may be because the two tasks are “simply different ‘kinds’ of memory errors that rely on different underlying mechanisms” (Ost et al., 2013, p. 5).

The results of the CFA indicated that a two-factor model, which comprised factors corresponding to false memory performance on the DRM and misinformation tasks, accounted for the data better than a one-factor model. However, the dissociation between the two factors may partially be explained by method variance. Furthermore, the analyses were exploratory since the variables associated with two of the tasks were excluded from the models due to poor psychometric properties. To account for this, the measurement models were modified and were not consistent with our original hypotheses. As such, an independent replication of the current results would strengthen confidence in the current findings.

The current findings provide evidence that there may be different underlying mechanisms responsible for the memory errors on each task (due to low correlation among the variables and the lack of convergent validity demonstrated by the DRM and misinformation variables). However, the current study did not directly assess which mechanisms best explain performance on each task. It is also worth noting that more than one mechanism may be necessary to explain each memory error and that there may be a common mechanism or framework that contributes to each (e.g., the source monitoring framework).

Strengths and Limitations

Although previous studies (e.g., Lövdén, 2003; Ost et al., 2013, Platt et al., 1998) have utilized multiple measures of false memory, the current study is the first to include four different measures that have been described as measuring false memory. An additional strength of the current study was the use of a red herring (see Laney et al., 2008) to distract participants from discovering the true purpose of the study. When participants were asked to generate an explanation describing the purpose of the study, only nine participants (8.00%) spontaneously mentioned something related to false memories or memory errors. However, close to 50% of participants selected false memory to be the true purpose of the study when given a forced-choice assessment. This,

in turn, may suggest that participants could have been aware of the phenomenon being studied, but may not have been able to provide a verbal description of the purpose of the study. Despite this, analyses showed that there were generally no differences between participants who had possible awareness of the study rationale and those that did not.

Despite the strengths of the current study, there are limitations. Notably, the study only had a moderately sized sample consisting mainly of college students ($M_{age} = 21.35$), which limits the generalizability of the findings. The use of a selective (“highly functioning”) sample such as college students may contribute to reduced variability in the false memory tasks. Therefore, it is possible that evidence of convergent validity would be stronger when the sample comprises participants with increased variability in performance. Latent variables are, theoretically, perfectly reliable and thus are well suited for assessing the relationship among the false memory measures that have low to moderate reliability. That said, the picture/word recall task in particular had low variance and low reliability and likely suffered from a floor effect due to the low rate of false memories that emerged. Likewise, the CEQ task exhibited several issues including very few lure trials which likely led to poor reliability, and two of the critical CEQ items were endorsed only by a small number of participants, hence reducing variability of the items and limiting their relationship with the other variables. Additionally, CEQ item 1, which asks participants about viewing cigarette items on television, may be problematic. Although these advertisements were banned in the United States in 1971, these individuals may have seen them in other countries (e.g., while on holiday), thus some individuals who endorsed this item, may, in fact, be reporting an accurate memory. However, because this item has been designated as a measure of false memory in Pezdek and Lam’s (2007) study, it was included in the current study. Future research should incorporate alternative items to further ascertain this measure’s relationship with other false memory tasks. In addition, the poor psychometric properties of the picture/word task and CEQ task suggest that false memory researchers should report the specifics of the reliability of their measures more often. Finally, although the literature on the relationship between these false memory paradigms and individual differences has been mixed (e.g., Bernstein et al., 2018; Hyman & Billings, 1998; Platt et al., 1998; Wilkinson & Hyman, 1998), it is possible that the specific tasks used in the current study may not adequately capture individual differences.

Conclusion

The results of the current study are consistent with previous research, which found little to no relationship between false memory in the DRM and the misinformation tasks (e.g., Ost et al., 2013). In addition, we also found that two other tasks used in the literature to assess false memory were only weakly related to one another, although this finding may be at least partially attributed to the poor psychometric properties of the measures. Furthermore, a CFA indicated that the DRM and misinformation variables were not adequately represented by a unitary factor. Thus, our findings suggest that tasks designed to measure false memory are likely measuring different types of memory errors. Future

research should examine whether different mechanisms are responsible for the different memory errors.

ACKNOWLEDGEMENTS

We gratefully acknowledge the following members of the Fordham University Memory & Aging Lab for their assistance in testing participants: Samantha Banjany, Olivia Cortellini, Emma Fetonti, Erin Hunt, Soulin Haque, and Maria Pleshkevich.

REFERENCES

- Arbuckle, J. L. (2016). Amos (Version 24.0) [Computer program]. Chicago: SPSS.
- Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability and validity of a dissociation scale. *Journal of Nervous and Mental Disease, 174*, 727–735. doi: 10.1097/00005053-198612000-00004
- Bernstein, D. M., Scoboria, A., Desjarlais, L., & Soucie, K. (2018). "False memory" is a linguistic convenience. *Psychology of Consciousness: Theory, Research, and Practice, 5*, 161–179. doi: 10.1037/cns0000148
- Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*, 164–169. doi: 10.1111/1467-8721.00192
- Braun, K.A., Ellis, R., & Loftus, E.F. (2002). Make my memory: How advertising can change our memories of the past. *Psychology and Marketing, 19*, 1–23. doi: 10.1002/mar.1000
- Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review, 11*, 921–925. doi:10.3758/BF03196722
- Calvillo, D. P., & Parong, J. A. (2016). The misinformation effect is unrelated to the DRM effect with and without a DRM warning. *Memory, 24*, 324–333. doi: 10.1080/09658211.2015.1005633
- DePrince, A. P., Allard, C. B., Oh, H., & Freyd, J. J. (2004). What's in a name for memory errors? Implications and ethical issues arising from the use of the term "false memory" for errors in memory for details. *Ethics & Behavior, 14*, 201–233. doi: 10.1207/s15327019eb1403_1
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*, 833–848. doi: 10.3758/MC.38.7.833
- Hyman, I. E., & Billings Jr, F. (1998). Individual differences and the creation of false childhood memories. *Memory, 6*, 1–20. doi: 10.1080/741941598
- Laney, C., Kaasa, S. O., Morris, E. K., Berkowitz, S. R., Bernstein, D. M., & Loftus, E. F. (2008). The Red Herring technique: A methodological response to the problem of demand characteristics. *Psychological Research, 72*, 362–375. doi: 10.1007/s00426-007-0122-6
- Leding, J. K. (2012). False memories and persuasion strategies. *Review of General Psychology, 16*, 256–286. doi: 10.1037/a0027700
- Lindsay, D. S., Wade, K., Hunter, M., & Read, J. D. (2004). Adults' memories of childhood: Affect, knowing, and remembering. *Memory, 12*, 27–43. doi: 10.1080/09658210244000243
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19–31. doi: 10.1037/0278-7393.4.1.19
- Lövdén, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49*, 268–283. doi: 10.1016/S0749-596X(03)00069-X
- Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17–24. doi: 10.1111/j.2044-8295.1973.tb01322.x
- Mayer, J. D., & Gaschke, Y. N. (1988). The experience and meta-experience of mood. *Journal of Personality and Social Psychology, 55*, 102–111. doi: 10.1037/0022-3514.55.1.102
- Mazzoni, G. (2002). Naturally occurring and suggestion-dependent memory distortions; The convergence of disparate research traditions. *European Psychologist, 7*, 17–30. doi: 10.1027//1016-9040.7.1.17
- Monds, L. A., Paterson, H. M., & Kemp, R. I. (2016). Do emotional stimuli enhance or impede recall relative to neutral stimuli? An investigation of two "false memory" tasks. *Memory, 25*, 945–952. doi: 10.1080/09658211.2016.1237653
- Nichols, R. M., & Loftus, E. F. (2019). Who is susceptible in three false memory tasks? *Memory, 27*, 962–984. doi: 10.1080/09658211.2019.1611862
- Ost, J., Blank, H., Davies, J., Jones, G., Lambert, K., & Salmon, K. (2013). False memory ≠ false memory: DRM errors are unrelated to the misinformation effect. *PLoS One, 8*, 1–6. doi:10.1371/journal.pone.0057939
- Otgaar, H., & Candel, I. (2011). Children's false memories: Different false memory paradigms reveal different results. *Psychology, Crime & Law, 17*, 513–528. doi: 10.1080/10683160903373392
- Otgaar, H., Verschuere, B., Meijer, E. H., & van Oorsouw, K. (2012). The origin of children's implanted false memories: Memory traces or compliance? *Acta Psychologica, 139*, 397–403. doi: 10.1016/j.actpsy.2012.01.002
- Patihis, L., Frenda, S. J., & Loftus, E. F. (2018). False memory tasks do not reliably predict other false memories. *Psychology of Consciousness: Theory, Research, and Practice, 5*, 140–160. doi: 10.1037/cns0000147
- Pezdek, K., & Lam, S. (2007). What research paradigms have cognitive psychologists used to study "false memory," and what are the implications of these choices? *Consciousness and Cognition, 16*, 2–17. doi: 10.1016/j.concog.2005.06.006
- Platt, R. D., Lacey, S. C., Lobst, A. D., & Finkelman, D. (1998). Absorption, dissociation, and fantasy-proneness as predictors of memory distortion in autobiographical and laboratory-generated memories. *Applied Cognitive Psychology, 12*,

- S77–S89. doi: 10.1002/(SICI)1099-0720(199812)12:7<S77::AID-ACP601>3.0.CO;2-C
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401. doi: 10.1177/014662167700100306
- Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9*, 95–123. doi: 10.1016/S1041-6080(97)90002-9
- Roberts, P. (2002). Vulnerability to false memory: The effects of stress, imagery, trait anxiety, and depression. *Current Psychology, 21*, 240–252. doi: 10.1007/s12144-002-1016-9
- Roediger, H. L., & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9*, 123–127. doi: 10.1111/1467-8721.00075
- Salthouse, T. A., & Siedlecki, K. L. (2007). An individual difference analysis of false recognition. *The American Journal of Psychology, 120*, 429–458. doi: 10.2307/20445413
- Schacter, D. L., Guerin, S. A., & St Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences, 15*, 467–474. doi: 10.1016/j.tics.2011.08.004
- Siedlecki, K. L., Stern, Y., Reuben, A., Sacco, R. L., Elkind, M. S., & Wright, C. B. (2009). Construct validity of cognitive reserve in a multiethnic cohort: The Northern Manhattan Study. *Journal of the International Neuropsychological Society, 15*, 558–569. doi: 10.1017/S1355617709090857
- Spielberger, C. D., Gorusch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*, 494–500. doi: 10.3758/BF03211543
- Takarangi, M. K., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: The development of a new stimulus set. *Applied Cognitive Psychology, 20*, 583–590. doi: 10.1002/acp.1209
- Wilkinson, C., & Hyman, I. E. (1998). Individual differences related to two types of memory errors: word lists may not generalize to autobiographical memory. *Applied Cognitive Psychology, 12*, 29–46. doi: 10.1002/(SICI)1099-0720(199812)12:73.0.CO;2-R
- Wilson, S. C., & Barber, T. X. (1978). The Creative Imagination Scale as a measure of hypnotic responsiveness: Applications to experimental and clinical hypnosis. *American Journal of Clinical Hypnosis, 20*, 235–249. doi: 10.1080/00029157.1978.10403940
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*, 913–934. doi: 10.1177/0013164413495237
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., & Dong, Q. (2013). The relationship between DRM and misinformation false memories. *Memory & Cognition, 41*, 832–838. doi:10.3758/s13421-013-0300-2

RECEIVED 05.07.2019 | ACCEPTED 20.10.2019

APPENDIX

TABLE A1.

Study Procedure and Task Description in Order of Presentation

Task name (in order of presentation)	Description
1. Mood Assessment 1 (Mayer & Gaschke, 1988).	A 1-item mood assessment derived from the Brief Mood Introspection Scale was administered on five occasions and asked participants to rate their current mood on a scale from 1-5.
2. DRM Recall (Stadler, Roediger, & McDermott, 1999).	Participants presented with eight word lists. Following administration of each list, participants completed a recall task by writing down all of the words they could remember in an answer booklet.
3. DRM Recognition Task (Salthouse & Siedlecki, 2007).	Participants were asked to read a list of words and report whether the item was new or old, and to report their confidence level on a scale of 1-5.
4. State Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983).	The trait version of the STAI was administered in 20 statements to examine feelings of anxiety.
5. Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973).	Participants were asked to visualize four different scenes and were asked to rate the vividness of the mental image on a 5-point Likert scale.
6. Mood Assessment 2	
7. Picture/Word Recall Task (Phase I; Roberts, 2002).	Participants were presented with 25 pictures and 25 words on a screen and were instructed to keep a mental tally of the number of inanimate and animate objects. Once the presentation concluded, participants were asked to report their tally.
8. Childhood Experiences Questionnaire (CEQ; Lindsay, Wade, Hunter, & Rad, 2004)	Participants read 32 statements regarding their childhood experiences, with four of these statements considered highly unlikely/impossible events and were asked to classify the strength of the memory on a scale of 1-5.
9. Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977)	A 20-item measure used to assess depressive symptoms on a 4-point Likert Scale.
10. Dissociative Experiences Scale (DES; Bernstein & Putnam, 1986).	A 28-item questionnaire assessing dissociation that occurs in daily experiences and the frequency of those experiences.
11. Mood Assessment 3	
12. Picture/Word Recall Task (Phase II)	After a 25-minute delay, participants were asked to recall of the pictures and words presented in phase I, then they were asked to make a remember/know judgement for each word recalled.
13. Mood Assessment 4	
14. Misinformation Paradigm (Phases I and II; Takarangi, Parker, & Garry, 2006).	Participants were shown a video on a computer screen about an electrician entering a client's home and stealing items. Then, participants were asked to read a passage with four items manipulated to introduce misleading post-even information.
15. Creative Imagination Scale (CIS; Wilson & Barber, 1978)	Three passages were read aloud and participants were asked to imagine the situation. Following this, participants were asked to rate each of the experiences in terms of how similar they felt to actual experiences.
16. Misinformation Paradigm (Phase III)	Participants were given a 20-item forced choice recognition task regarding the details of the video that was watched in phase I.
17. Mood Assessment 5	
18. Demographic Questionnaire	Participants were asked to provide basic demographic information such as age, gender, and education.
19. Awareness Questionnaire	Participants were asked to report what they believed the purpose of the study was in an open-ended question, then they were administered a multiple choice question to indicate which of the five alternatives they believed best described the purpose of the study.