False Memory for Words in Noise: An At-Home Deese-Roediger-McDermott (DRM) Experiment Across Adulthood

Rebecca L. Wagner1,2, Bethany A. Lyon1,*, and Angela AuBuchon2,*
1Department of Psychology, University of Nebraska at Omaha
2Boys Town Medical Research Hospital

ABSTRACT. The Deese-Roediger-McDermott paradigm is an experimental manipulation that induces false memory creation (Deese, 1959; Roediger & McDermott, 1995). This study presented auditory DRM lists either in silence or in background noise to both young and older adult samples. Background noise significantly reduced participants’ recall accuracy, \( F(1, 93) = 14.14, p < .001, \eta_p^2 = .13 \), and false memory production, \( F(1, 93) = 4.02, p = .05, \eta_p^2 = .04 \), but there were no significant differences between age groups. Moreover, older adults’ self-reported measures of cognitive load correlated with their accuracy, \( r(47) = -.39 \). No such relationship was observed in the younger adult group suggesting that, although both groups reached similar levels of performance, they likely approached the task differently. These findings also highlight the need for further examination into how individual differences mediate memory performance in the DRM, especially in older adults.

Keywords: cognitive, memory, older adults, free recall, background noise

Memory is fallible. Knowing the conditions under which memory fails highlights points for intervention and contributes to theory of underlying processes that support memory. Typically, memory failures are associated with information that is presented but later forgotten. Another type of memory failure is when information is seemingly “remembered” but was never presented. These occurrences, known as false memories, are difficult to confirm in daily life because they depend on reliable documentation of the presented information. However, false memories are easily observed in laboratory experiments where the presented information is known.

One method commonly used to elicit false memories is the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). The DRM uses lists of 12 to 15 highly related words that fall within a semantic network (e.g., nap, rest, lay, dream, awake). The words are typically presented individually with a free recall or recognition test completed immediately afterward. Most importantly, each presented list excludes a “critical lure” word (e.g., sleep), which is closely related to the listed words (e.g., nap, rest, lay). Recall of the critical lure serves as a marker for a false memory provoked by the DRM paradigm (Deese, 1959; Roediger & McDermott, 1995). Many different factors influence the rate of false memory production in the DRM, such as aging, including associated changes in processing approaches, list structure, and more.

Aging is associated with decreased recall of presented words and an increase in inclusion of critical lures and other list intrusions (Watson et al., 2005). Changes to both cognition and sensory systems appear to contribute to older adults’ increased susceptibility to false memories. Declines in cognitive functioning, particularly in attention allocation abilities, increase the task difficulty for
older adults (Kahneman, 1973). When the DRM list is presented aurally, task difficulty is further increased for individuals with reduced hearing acuity, common among older adult populations.

**The Role of Suppression Mechanisms in Preventing False Memories**

Production of false memories in the DRM arises from the inability to suppress inappropriately activated items, the critical lures, during recall of presented lists. Attention and source monitoring, a set of processes that aid in differentiating between original sources of information (Hashtroudi et al., 1990; Johnson et al., 1993), are integral for suppressing these errors in activation elicited by the DRM’s design. Older adults’ reduced access to attentional resources can cause source-monitoring failures during recall, typically resulting in an overall higher recall of words than their younger counterparts, but more of the words are critical lures, outside list intrusions, and repeated words (Dywan & Jacoby, 1990; Watson et al., 2005).

In memory, related words, phrases, and concepts are linked, allowing an entire network of related concepts to be activated by a single word or idea. Typically, this spreading activation is beneficial during memory recollection, because singular memory cues can activate entire networks of information, making that information more readily accessible, rather than needing specific cues for each individual memory (Anderson, 1983). However, the DRM relies on this spreading activation bias in memory to activate critical lures during list recall, creating a false memory which leads to inaccurate recollections (Pierce et al., 2005). Adequate attentional control and source monitoring processes are then critical for inhibition of spreading activation caused by the DRM list presentation (Meade et al., 2006; Watson et al., 2005).

The effects of spreading activation and need for accurate source monitoring on DRM list-memory are amplified by the changes in processing approaches that occur with aging. Younger adults tend to rely on item-based, or verbatim, processing. In other words, each list item is processed and encoded on its own (Tun et al., 1998). This method is known to require more cognitive resources and, in general, be more effortful to complete (Brainerd & Reyna, 2005). Older adults, on the other hand, tend to rely more on gist-based processing, focusing on commonalities among items rather than the items themselves (Tun et al., 1998). Differences in gist processing and item-based processing align with underpinnings of the fuzzy trace theory (Reyna & Brainerd, 1998). This theory posits that individuals can create two representations of the same information, one that is exact to the information presented, and one that is the gist, or general idea of the information presented. Reliance on gist processing in a DRM paradigm leads participants to generate false memories, as the critical lure matches the gist of the entire list. Older adults’ preference for gist-based processing, coupled with the noted attentional demands of a DRM task, increase the likelihood that a critical lure will be recalled.

**List Structure Influences False Memory Production**

Accompanying age-related changes that facilitate false memory production, the structure of the DRM lists can further increase the likelihood that a critical lure will be recalled. Each DRM list consists of words that are highly related to the unpresented critical lure; however, some lists have greater associative connections to this word compared to others. When this association, called the backward associative strength (BAS), is stronger, the rate of false memory production tends to be higher (Knott et al., 2012).

**Background Noise Imposes Further Cognitive Demand**

Presenting auditory memoranda in background noise further interferes with processing and impairs recall by imposing additional cognitive demands on adult listeners (Hintzman, 1988; Peelle, 2018). Background noise reduces perceptual information available to a listener, increasing the chances that presented information is misidentified and stored inaccurately in memory, which would manifest as a memory error (Peelle, 2018). To decrease the likelihood of an error, attentional resources are diverted away from semantically processing the words to perceptually processing the individual sounds of the words. This results in better identification of list words but decreases the total number of words recalled. Additionally, perceptual processing limits the amount of spreading activation that occurs, which decreases the number of critical lures and list intrusions present in recall (Kjellberg et al., 2008; Marsh et al., 2015; McCoy et al., 2005).

**Individual Differences in Cognitive Load**

Cognitive load refers to the combination of cognitive abilities and mental effort that individuals allocate toward a task, as well as one’s metacognitive awareness of the demands of that task (Tomporowski, 2002, 2003). Cognitive abilities, including attention allocation, perceptual
processing, memory, and perceptions of task demands, all vary with individuals. Mental effort, or the effort an individual puts toward a task, also varies between individuals, as it has been shown to be affected by past experience, conceptions surrounding skill level and ability, motivation, and age (Van Gerven et al., 2000). The National Aeronautics and Space Administration Raw Task Load Index (NASA RTLX) has been a validated tool to assess the subdivisions of cognitive load in both younger adults (Longo, 2018), older adults (Devos et al., 2020), and listening-in-noise tasks (Mackersie & Cones, 2011). Closer examination of individual differences in cognitive load using the DRM, with and without background noise, is useful in teasing apart the complex nature of false memories, their formation, and why some are more susceptible to their formation than others.

Current Study
For the current study, we aimed to assess recall accuracy and false memory production in adults on an at-home, auditory DRM task. Based on previous literature, we expected that adding noise to an auditory DRM task would increase the cognitive load of the task, decreasing both recall accuracy and the number of false memories produced. We anticipated this change in performance to be more pronounced in older adults due to changes in hearing ability, noted shifts in processing approaches, and additional cognitive loads that are introduced by background noise.

Method
Participants were anonymously recruited and remotely tested using web-based platforms prolific.co and gorilla.sc. Prolific.co was utilized for completely anonymous, online participant recruitment in which participants were prescreened to meet study inclusion criteria by the platform. Following their completion of the study, participants were automatically paid by the platform. In the event that participants needed to contact a researcher, they could use the platform’s messaging system, which maintained their anonymity. Gorilla.sc, on the other hand, was used to administer all experiment procedures completely online via a web browser on participants’ personal computers. The link, generated on gorilla.sc, was provided to prolific.co, allowing participants to be seamlessly directed between the two platforms.

In the current study, participants received DRM lists presented auditorily over headphones either in silence or embedded in multitalker background noise. Prior to hearing the noisy lists, participants completed a 15-word speech perception task that assessed their ability to hear words in noise. After completing both blocks, participants were asked to self-report their conceptions about the task’s demands and their ability to meet those demands using the NASA RTLX. Treatment of all participants met the guidelines set forth by the American Psychiatric Association and a university-based institutional review board.

Participants
One hundred participants were screened and recruited anonymously using prolific.co. This platform uses participant self-report questionnaires to allow researchers to screen their participants based on a variety of criteria. Our young adult participants all reported (a) being between 19 and 29 years of age ($M = 23.25; SD = 3.05$), (b) being from, and currently living in, the United States (U.S.), (c) learning and speaking English as their first language, (d) and having perceived normal or corrected to normal vision, no known hearing difficulties, and no known long-term health conditions or disabilities. Our older adult participants all reported being between ages 65 and 75 ($M = 68.78; SD = 2.87$) and met Criteria 2 through 4 mentioned above. Five participants, four young adults and one older adult, reported writing words down or receiving help from someone else on the task and were excluded. The final analyses included 46 younger adults and 49 older adults (see Table 1 for reported
participant demographics). All participants passed the speech perception task \((M = 0.91, SD = 0.08)\), the attention checks within the study, and were compensated $4 through prolific.co.

**Materials**

**DRM Lists**

Twenty-four DRM lists were recorded from Roediger and McDermott (1995). These were recorded in a female voice using an anechoic chamber. Using Audacity for Windows (Version 2.4.2, Audacity), the word lists were created to present a word every 2 seconds in the same order as Roediger and McDermott (1995). Lists were leveled to present the words between 60 and 70 dB with the computer set at 50% volume. Of the 24 lists recorded, we utilized 23 of them throughout the task. King and river lists were used as practice lists because they included proper nouns (i.e., England, George, and Mississippi). The girl list was used in the speech perception task. The spider list was excluded from presentation, as it included words such as “tarantula” and “arachnid,” which are commonly misspelled. The remaining 20 lists were used during the memory conditions of the task.

**Background Noise**

In order to mimic a natural environment, cafeteria ambience noise recorded using a TASCAM DR-40 portable recorder at 48kHz 24bit (Stereo) was selected from freesound.org (https://freesound.org/people/TaXMaNFoReVeR/sounds/325438/). For presentation during the lists, the first 40 seconds of the sound clip was extracted and dampened to 45 to 55 dB, roughly 10 to 15 dB quieter than the word presentation decibel level. The sound presentation was then combined with the word lists, maintaining the decibel difference during presentation, even across different participant-selected volumes. All sound files reported are available on OSF (https://osf.io/gu7aw/).

**NASA Raw Task Load Index**

Our NASA RTLX procedure was adopted from the NASA TLX pen-and-paper version available online (https://humansystems.arc.nasa.gov/groups/TLX/tlxpaperpencil.php). Participants were asked to use a slider to rate the following categories: mental demand, physical demand, temporal demand, performance, effort, and frustration. Mental demand asks how mentally demanding participants interpreted the task to be. Physical demand asks how physically demanding participants perceived the task to be. Temporal demand asks how hurried or rushed the pace of the task seemed. Performance asks participants to rate how successful they felt they were at accomplishing what was asked of them during the task. Effort asks participants how hard they had to work in order to accomplish the level of performance they achieved. Lastly, frustration asks participants the level to which they felt “insecure, discouraged, irritated, stressed, or annoyed” during the task. Each response, aside from performance, was ranked from **very low** to **very high**. Performance was rated from **perfect to failure**.

The original NASA TLX procedure proposed by Hart and Staveland (1988) included both the aforementioned rating scale task and a pair-wise rating task where participants are presented two of the six item titles (i.e., mental demand, physical demand, temporal demand, performance, effort, frustration) and are asked to choose which was more important to them when completing the task (Hart & Staveland, 1988). A weighted score would then be calculated from both the rating item responses and the pair-wise rating task. The NASA RTLX procedure only collects participants’ item raw scores, which can then be averaged or analyzed individually (Hart, 2006). Longo (2018) reported an alpha reliability coefficient of .65 when administering the NASA RTLX with young adults. With older adults, Devos and colleagues (2020) reported an interclass correlation coefficients (ICCs) between .71 and .81, indicating good reliability for the NASA RTLX with older adults. The present study reported an alpha reliability coefficient of .54 for the NASA RTLX across both age groups.

**Procedures**

All experimental procedures were conducted via gorilla.sc (Anwyl-Irvine et al., 2019). Prior to starting the experiment, participants completed a headphone-check task adopted from Parker and colleagues (2021) in which participants had to determine which of three tones was the quietest. They had two attempts to pass four of six trials. If they were unable to pass the headphone check, they were disenrolled from the study. After passing the headphone check, participants were randomly assigned to start with either the silent list block or the noisy list block to minimize practice effects. All participants received both block types.

At the conclusion of the experiment, participants completed the NASA RTLX and were asked if they noticed anything about the words and the list structures. No participants reported any prior knowledge regarding the DRM.
Experimental Tasks
Participants completed two blocks of trials each consisting of 12 DRM lists: two practice and 10 trial lists. One block presented the lists in silence while the other presented lists embedded in the background noise, counterbalanced across participants. Trial lists were randomly selected for each block by the experimental program and only appeared once for each participant. After listening to one DRM list, participants were prompted to type the words they recalled from the list. After finishing their recall task, participants advanced to the next list until all 10 trial lists were completed for the block. When using these measures, we reported an alpha reliability coefficient of .70 for the silent list blocks and .71 for the noise list blocks.

Prior to beginning the noisy block, participants completed the speech perception task. All 15 words from the girl list were presented individually within background noise. Following each word, the free response box would appear, prompting participants to enter the word they heard into the box.

Analysis Plan
Data cleaning to correct misspellings of typed responses was completed prior to calculating accuracy and false memory rates. Three raters who were blind to the experiment procedures and expectations analyzed the spelling of response. If all three raters agreed on what word the participant was attempting to spell, the spelling was changed and the response was included in the analyses. Many of the corrected responses were common spelling mistakes or homophones (e.g., switching “ie” and “ei” or “course” and “coarse”). All spelling corrections were held constant across both groups.

Accuracy and false memory were calculated based on whether the word or critical lure, respectively, were present in participants’ recalled responses. Averaged accuracy was calculated from the number of presented list words recalled per list. Average rates of false memory were calculated from the number of critical lures present in the recall responses. BAS values for the 20 DRM lists were taken from Roediger and colleagues (2001). NASA RTLX item scores were left in their raw form. RTLX total load scores were calculated by averaging participants’ responses, a process that produces more or equally sensitive results as the original weighting procedure (Bittner et al., 1989; Hendy et al., 1993).

All mixed ANOVA analyses were conducted using the rstatix R package (version 0.7.0; Kassambara, 2021) in RStudio (version 1.4.1717; 2021). The anova_test function was used to calculate mixed methods ANOVAs for accuracy and false memories across both age groups (young adult, older adult) and task type (silent, noise). Pearson correlation analyses were conducted using the stats R package. The cor.test function was used to perform correlations between the NASA RTLX self-report scores, accuracy, and false memory rates between the age groups.

Results
We conducted a 2 (age group) x 2 (block type) mixed methods ANOVA for recall accuracy. The main effect of age group for recall accuracy, $F(1, 93) = 0.84, p = .36, \eta_p^2 = .01$, was not significant. A main effect of block type was found, $F(1, 93) = 14.14, p < .001, \eta_p^2 = .13$, with better accuracy in the silent block, as anticipated. Younger adults performed slightly better in silence ($M = 7.47$, $SD = 2.45$) than in noise ($M = 7.16$, $SD = 2.55$). The same was found for older adults, who performed better in silence ($M = 7.92$, $SD = 2.56$) than in noise ($M = 7.41$, $SD = 2.44$). The interaction was not significant, $F(1, 93) = 0.87, p = .35, \eta_p^2 = .01$ (see Figure 1).

We also conducted a 2 (age group) x 2 (block type) mixed methods ANOVA for false memories measured as the number of critical lures recalled. The findings for accuracy were mirrored in the false memory ANOVA pattern: the main effect for age group, $F(1, 93) = 0.11, p = .74, \eta_p^2 = .00$, was not significant, but the main effect for block type was significant, $F(1, 93) = 4.02 p = .05, \eta_p^2 = .04$. 

FIGURE 1

Group Accuracy Between Block Conditions

Note. Accuracy scores were calculated based on whether or not a presented word was present in participants’ recall. Because of this, participants could have up to 15 items correctly recalled.
Within groups, between the silent and noise blocks, younger adults recalled more critical lures during the silence ($M = 0.30, SD = 0.46$) than noise ($M = 0.22, SD = 0.42$). Older adults had similar degrees of false memories across silence ($M = 0.28, SD = 0.45$) and noise ($M = 0.27, SD = 0.45$). The interaction was not significant, $F(1, 93) = 2.61, p = .11, \eta^2_p = .03$ (see Figure 2).

Similar to previous research, positive correlations between each list’s BAS and the recall of the list’s critical lure were significant. Young adult critical lure recall was positively correlated with the list BAS, $r(44) = .42, p = .003$. In addition, older adult recollection of the critical lure was positively correlated with the list BAS, $r(47) = .53, p < .001$ (see Figure 3).

**Exploratory Analysis**

It was important for us to tease apart the individual differences in cognitive load for the DRM task, as this differs across individuals, and possibly across age groups. To examine participants’ cognitive load on the DRM task more closely, we conducted a correlational analysis between the NASA RTLX total load score, the item scores (i.e., mental demand, physical demand, temporal demand, performance, effort, frustration), and group-level accuracy and false memory rates. Because we only administered the NASA RTLX once at the end of the task, we collapsed across the sound condition blocks to perform the correlational analyses. Accuracy and false memory relations with the total load score and physical demand, performance, frustration scores reached significance, primarily with older adults. Further elaboration of the total load score, physical demand, performance, and frustration correlations are outlined below, but all correlations and their significance are reported in Table 2.

**Total Cognitive Load**

Averaged total load scores were negatively correlated with older adult accuracy, $r(47) = -.39, p = .01$ (see Figure 4A), indicating that, as older adults reported higher levels of cognitive load through their reports, their accuracy for the presented words decreased. Older adult rates of false memory were not significantly correlated with their perceived cognitive load, although this was nearing significance, $r(47) = .27, p = .06$ (see Figure 4B). Young adult averaged cognitive load accuracy, $r(44) = .02, p = .91$, and false memory, $r(44) = .06, p = .69$, correlations were not significant (see Figures 4A & 4B).

---

**TABLE 2**

<table>
<thead>
<tr>
<th></th>
<th>Young Adult Accuracy</th>
<th>Young Adult False Alarms</th>
<th>Older Adult Accuracy</th>
<th>Older Adult False Alarms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Cognitive Load</td>
<td>-.02</td>
<td>-.06</td>
<td>-.39*</td>
<td>-.27*</td>
</tr>
<tr>
<td>Mental Demand</td>
<td>.13</td>
<td>-.07</td>
<td>.33*</td>
<td>-.01</td>
</tr>
<tr>
<td>Physical Demand</td>
<td>.04</td>
<td>-.03</td>
<td>-.42*</td>
<td>.19</td>
</tr>
<tr>
<td>Temporal Demand</td>
<td>.17</td>
<td>-.24</td>
<td>-.17</td>
<td>-.01</td>
</tr>
<tr>
<td>Performance</td>
<td>-.32*</td>
<td>.20</td>
<td>-.49*</td>
<td>.38*</td>
</tr>
<tr>
<td>Effort</td>
<td>.16</td>
<td>.05</td>
<td>.18</td>
<td>-.17</td>
</tr>
<tr>
<td>Frustration</td>
<td>-.02</td>
<td>-.16</td>
<td>-.30</td>
<td>.32*</td>
</tr>
</tbody>
</table>

Note. Removal of one outlier removed the significance and changed the correlation to $r(46) = .18, p = .22$. *significance < .05, **significance < .001.
**Physical Demand**
Older adult physical demand subscores were negatively correlated with their performance, $r(47) = -0.43$, $p = 0.002$ (see Figure 5A), implying that older adults who reported the task requiring less physical effort to complete had greater accuracy scores than older adults who reported that the task required more physical effort. Older adult false memory rates, $r(47) = 0.19$, $p = 0.19$ (see Figure 5B), along with young adult accuracy, $r(44) = 0.04$, $p = 0.80$, and false memory rates, $r(44) = 0.03$, $p = 0.82$, correlations were also not significant (see Figures 5A & 5B).

**Performance**
Performance subscores were negatively correlated with older adult accuracy, $r(47) = -0.49$, $p < 0.001$, and younger adult accuracy, $r(44) = -0.32$, $p = 0.03$ (see Figure 5A), indicating that, regardless of age, participants who reported feeling more successful at completing what the task asked of them (i.e., rating scores closer to 0, or perfect) were more accurate in their list recall. However, performance subscores

---

**FIGURE 4**

**NASA RTLX Averaged Total Cognitive Load Scores Correlated With Group Accuracy (A) and Group False Memory Rates (B)**

A

![Graph A](image1.png)

B

![Graph B](image2.png)

**FIGURE 5**

**NASA RTLX Physical Demand Subscores Correlated With Group Accuracy (A) and Group False Memory Rates (B)**

A

![Graph A](image3.png)

B

![Graph B](image4.png)
only correlated with false memory rates for older adults, $r(47) = .38, p = .01$ (see Figure 6B). Therefore, older adults who thought poorly about their performance (i.e., reported scores closer to 100, or failure) on the task had an increased number of critical lures present in their recall of the presented lists. The correlation between false memory rates and performance subscores was not significant for younger adults, $r(44) = .20, p = .17$ (see Figure 6B).

**Frustration**
The frustration item assessed participants’ feelings of insecurity, discouragement, irritation, stress, or annoyance during the task. For older adults, there was a significant negative correlation between frustration subscores and accuracy, $r(47) = –.30, p = .04$, meaning that, as older adults reported greater levels of these feelings due to the task’s demands, their recall accuracy for the presented list words decreased (see Figure 7A). There was also a significant positive correlation between frustration scores and false memory rates in older adults, $r(47) = .32, p = .03$, indicating that, when older adult participants reported greater levels of the feelings covered under the frustration item, the appearance of critical lures in the participants’ list recall increased (see Figure 7B). In the younger adult group, both accuracy, $r(44) = –.02, p = .88,$
and false memory, $r(44) = -0.16, p = .30$, correlations were not significant (see Figures 7A & 7B).

**Discussion**

For the current study, we aimed to examine age-related differences in accuracy and false memory rates between younger and older adults on a remote, auditory DRM task. We expected to amplify these differences with the addition of background noise, which would increase the cognitive load of the task. This was expected to affect older adults more severely because of changes in hearing acuity, default processing type, and the ability of older adults to meet the cognitive load of the task. These manipulations were designed to help us better understand adult susceptibility to false memory production in everyday listening environments.

We did observe a significant main effect between noise and silent conditions on both participants’ accuracy as well as the number of false memories they produced. However, this effect was similar for both younger and older adults. Background noise, even when it is lower than conversational speech levels, appeared to decrease the number of false memories created through the DRM’s associative, semantic spreading activation. However, how the DRM phenomenon arises in noise conditions, whether through affecting perceptual access to the presented words, individual differences, or some interaction of the two, still remains unclear.

In addition, we found a significant positive correlation between the BAS of each DRM list and the rate of false memory production on that list across both age groups, further supporting the importance of strong inner list word association to produce false memories in the DRM. However, it is important to highlight that these positive correlations were similar for both our younger and older adult age groups, indicating that, in our sample, the strength of the BAS affected false memory production recall in both groups in a similar way.

Given previous research and understanding of memory changes with age, it is intriguing that we did not find significant differences between age groups. When comparing younger and older adults, equivalent performance on any memory task is atypical. A post-hoc power analysis determined a weak power level for the current study (1-$\beta = .67$), however, Norman and Schacter (1997), which provided the basis of the current study, detected age-group differences in a free recall DRM task with only 24 participants in each group. Given that the current study doubled that of Norman and Schacter, it is unlikely that we were underpowered to find age-group differences.

Comparable performance in both accuracy and false memory rates between our younger and older adult participants may be a driving force behind our lower power level and may indicate differences in our adult samples compared to those of previous literature. Older adult participants in this study were active on prolific.co, which may imply they use the internet more frequently, engage with studies more often, and thus may differ from older adults who generally participate in research in-person or from the general older adult population in the United States. Further research is needed to better understand whether, or what, age-related differences may affect how people approach listening to and remembering information presented noise that more closely resembles our everyday environment.

Although our older adult population might have differed from the typical older adult population of the United States, our exploratory analyses attempted to tease apart perceived differences in task demands that might have affected task performance. Through NASA RTLX self-reports, we found that self-conceptions surrounding our task, its demands, and participants’ own cognitive abilities (i.e., the cognitive load) related to their accuracy and susceptibility to false memory production on the DRM task, particularly for our older adults. These results may indicate differences in older adults’ perceptions of the cognitive load of the task at hand, causing differences in the cognitive resources allocated to the task and the mental effort they put toward accomplishing the task.

The exploratory results may also highlight differences in older adults’ level of cognitive self-efficacy, or their beliefs surrounding their own ability to reach certain levels of performance on cognitively demanding tasks (Seeman et al., 1996). Those who reported that the task required more physical demand from them, that their performance was inadequate, or that they were more frustrated with the task may have weaker cognitive self-efficacy. Weaker self-efficacy surrounding cognitive ability could also lead to older adult individuals putting less effort toward the task or prevent them from adequately meeting the task’s demands (Seeman et al., 1996). Looking forward, these perceptions of task demands, and conceptions of participants’ own cognitive abilities can provide direction for further understanding when there is a lack of age differences found in performance for...
memory-dependent tasks, similar to the findings of the current study.

Limitations
The at-home, independent testing removed some experimental and environmental control in our experiment. For example, because participants could set their own volume levels, we had to rely on relative signal-to-noise ratios without being able to set absolute listening levels. It is also possible that participants did not adhere to our directions (writing down words without disclosing it, etc.). Nevertheless, precautionary checks, such as requiring headphones to limit nonexperimental background noise, were implemented to catch these potential limitations brought about by remote testing procedures.

Aside from limitations in procedural control within the present study, our participant pool may also present an additional, unexpected limitation. Unlike previous, in-person administrations of the DRM with older adult participants, our study did not find age-related differences in memory accuracy or false memory production. This finding brings about additional questions in how our older adult sample was different than that of previous studies. For example, self-selecting into Prolific’s participant pool likely requires a certain level of comfort learning new technologies, which may be related to overall cognition. Or older adults on prolific may be practiced at memory tests if they have been exposed to other similar experiments on the platform. Unfortunately, due to the anonymity of our participant pool, it is unknown if there were specific demographic factors that prevented us from finding age-related differences in memory ability. Future studies that utilize remote testing of older adults on a DRM paradigm should include more in-depth demographic questionnaires to account for potential individual differences. Or older adults on Prolific may be practiced at memory tests if they have been exposed to other similar experiments on the platform.

In addition to the aforementioned limitations, the low reliability coefficient for the NASA RTLX exploratory analyses in the present study does bring a sense of hesitancy when interpreting the correlational results. Although our reliability for this measure was considered low, other studies with both younger and older adults have reported higher reliability coefficients (see Longo, 2018, and Devos et al., 2020, respectively), potentially highlighting the need for more standardized administration of the RTLX across studies, particularly those in which the participant completes the measure without the guidance of a present researcher.

Application and Future Directions
The present study highlights the benefits of including measures of individual difference in studies of memory and aging, particularly in studies on false memory. As stated above, older adults tested using asynchronous and completely remote procedures performed comparably to younger adults—a departure from previous findings in the false memory literature (see Norman & Schacter, 1997). Uncovering the individual characteristics that support accurate free recall without increasing false memory in older adult populations will provide guidance to promote healthy cognitive aging. Future studies, especially those that implement remote procedures, should take additional steps to account for potential individual differences, especially those that are brought about by online recruitment platforms like prolific.co.

In addition to including more in-depth demographic questionnaires to address potential sample differences, the present study also reinforced the usefulness of the NASA RTLX in understanding older adults’ perceptions of cognitive load during memory tasks. Although the reliability of this measure has fluctuated across studies and administration procedures, assessing cognitive load during memory tasks like a DRM paradigm is essential for understanding memory accuracy, particularly when testing older adults. NASA RTLX items, particularly performance and frustration, were sensitive to older adult free recall accuracy (see Table 2). Despite also completing the NASA RTLX, younger adults’ self-reports, aside from performance, were not correlated to their accuracy (see Table 2), suggesting that younger and older adults approached the tasks differently, despite reaching similar levels of memory performance. With these findings, it is apparent that measures of cognitive load should be an addition to any testing procedure, whether remote or in-person, where task demands can increase perceived cognitive load by a participant.

Conclusion
This study expands our understanding of how task demands intersect with individual differences to promote false memory formation. Although previous research has examined DRM list performance under broadband noise presentation in young adults (Marsh et al., 2015), the multitalker babble used here reflects a common source of background noise faced by both older and younger adults in everyday listening situations. This type of background noise decreases list recall accuracy,
which also decreases the number of false memories produced during recall, similar to what has been noted in previous literature (see Kjellberg et al., 2008; Marsh et al., 2015; McCoy et al., 2005). This study also makes a notable contribution to the DRM literature by introducing assessments of participants’ self-conceptions regarding the task and their performance on the task in relation to both accuracy and susceptibility to form false memories. These self-conceptions provide further insight into individual differences in recall accuracy and false memory production within a DRM task, particularly when participants are older adults. Future studies utilizing the DRM paradigm should consider including similar measures, in addition to measures of hearing thresholds, especially when testing older adults.

References


Author Note. Rebecca L. Wagner © https://orcid.org/0000-0003-1659-6016

Bethany A. Lyon © https://orcid.org/0000-0002-8165-3075

Angela AuBuchon © https://orcid.org/0000-0002-6746-1462

This project was funded by the Fall 2020 Psi Chi Undergraduate Research Grant. Creation of the stimuli...
was supported, in part, by the National Institute of General Medical Sciences, U54 GM115458, which funds the Great Plains IDeA-CTR Network.

We would like to thank Dr. Chris Stecker assistance with recording and processing stimuli. We would also like to thank Dr. Aryn Kamerer for guidance on administering and analyzing the NASA Task Load Index.

We have no conflicts of interest to disclose. Materials and data for this study can be found at https://osf.io/gu7aw/.

Correspondence concerning this article should be addressed to Rebecca L. Wagner at 402-269-0729. United States. Email: rebeccawagner.rfw@gmail.com
Turn Your Compassion Into a Career in Mental Health

Our online degree programs feature relationship-focused curriculum, innovative faculty, and lead to licensure.

alliant.edu/psychology
Call for Doctoral-Level Reviewers

*Psi Chi Journal* is seeking reviewers interested in providing constructive feedback on our authors’ original empirical research. To increase the journal’s scope and relevance, diverse people of varied racial, ethnic, gender identity, sexual orientation, religious, and social class backgrounds are welcomed. If you have a doctoral degree in psychology or a related field, please join us.

**TO SUPPORT OUR EXCEPTIONAL REVIEWERS**

- Our online portal allows you to submit specific subject areas that you feel comfortable with reviewing.
- At any time, you can set unavailability dates so that we will know when not to contact you.
- With each request to review, you will receive basic instructions and a template review form, which you can use in order to save you time during your review.

**TO BECOME A REVIEWER, EMAIL**

steve.rouse@psichi.org

---

**An Eye on Graduate School**

*Psi Chi’s* digital anthology brings together our very best advice about applying to graduate school—advice accumulated from 25+ experts in over 20+ years of *Eye on Psi Chi* magazine issues.

In nineteen chapters, this eBook will help you navigate the seven primary steps that are vital to your acceptance at the graduate program of your choice.

- Preparing for and Selecting a Graduate Program
- The Application Process
- Preparing for the GRE
- Soliciting Letters of Recommendation
- Writing the Personal Statement
- Interviewing
- Choosing a Program and Succeeding in Graduate School

**Download Today at Store.PsiChi.org**

Log in as a Psi Chi Member to receive a discount!

*Kindle Edition Available Now!*
Publish Your Research in *Psi Chi Journal*

Undergraduate, graduate, and faculty submissions are welcome year round. Only one author (either first author or coauthor) is required to be a Psi Chi member. All submissions are free. Reasons to submit include

- a unique, doctoral-level, peer-review process
- indexing in PsycINFO, EBSCO, and Crossref databases
- free access of all articles at psichi.org
- our efficient online submissions portal

View Submission Guidelines and submit your research at [www.psichi.org/?page=JN_Submissions](http://www.psichi.org/?page=JN_Submissions)

---

Become a Journal Reviewer

Doctoral-level faculty in psychology and related fields who are passionate about educating others on conducting and reporting quality empirical research are invited to become reviewers for *Psi Chi Journal*. Our editorial team is uniquely dedicated to mentorship and promoting professional development of our authors—Please join us!

To become a reviewer, visit [www.psichi.org/page/JN_BecomeAReviewer](http://www.psichi.org/page/JN_BecomeAReviewer)

---

Resources for Student Research

Looking for solid examples of student manuscripts and educational editorials about conducting psychological research? Download as many free articles to share in your classrooms as you would like.

Search past issues, or articles by subject area or author at [www.psichi.org/journal_past](http://www.psichi.org/journal_past)

---

Learn About Psi Chi

Psi Chi is the International Honor Society in Psychology. Membership is primarily open to undergraduates, graduate students, transfer students, full-time and part-time faculty members, and alumni.

See membership benefits and a link to apply at [www.psichi.org/page/member_benefits](http://www.psichi.org/page/member_benefits)