Reviewer Comments to Author (if any):  
    Reviewer: 1  
      
    Comments to the Author  
    The current study examined how different conceptualizations of relatedness influenced participants judgments of relatedness and their recall of word pairs. Participants were presented with paired associates that were selected from databases based on a broad range of associative relatedness, semantic relatedness, and thematic relatedness. In separate blocks, participants provided judgments of the three types of relatedness on a series of word pairs, and then completed a final cued-recall test for all items. Results showed that semantic relatedness had a strong predictive value in determining the judgments of relatedness participants gave, even when the judgments were based on a different type of relatedness (i.e., for semantic and thematic relatedness, in addition to associative relatedness). The analyses describe the interactive nature of the three measures of relatedness in predicting participants judgments of relatedness and cued recall performance.   
      
    It is difficult to discern a compelling motivation for the current research. The authors state that "the present study seeks to provide further insight by examining how different levels of associative overlap (measured in FSG), semantic overlap (feature overlap measured with COS), and thematic overlap (measured with LSA) affect cognitive tasks such as short term item retrieval and item relatedness judgments." (p. 8). But why is this important? For selecting stimuli for paired associate learning experiments? Or, as hinted at near the end of the discussion, is the idea to modify models of (semantic) memory? If so, this would be a more compelling motivation, but should be strengthened throughout.  
      
    If the original JAM data were based on FSG only, is there any reason to think that COS and LSA would have similar relationships (i.e., would replicate the same slope/intercept pattern)? Is the idea not that these different types of associations are in fact, distinct from one another? So why would we necessarily expect these other types of relatedness to be judged in the same way as associative strength? The paper should provide more explicit support for the prediction of replicating the JAM function with the other two measures. Tried to expand on this in the description of hypothesis 1 and added a sentence at the end of the last paragraph before the application to judgments section.  
      
    The hypotheses need to be reiterated in the results section, especially as this is how the sections of the results are organized. The analysis of Hypothesis 2 is especially confusing, and I believe there is an error near the end of this section on page 18: "Thus, at high levels of COS, FSG and LSA are complementary when predicting recall, increasing…" (FIXED) This analysis is not examining recall, but only the relatedness judgments (when type of judgment is controlled). Please clarify the gist of this complex analysis, and focus on what is meant by controlling the type of judgment (FIXED). That is, is the idea to determine exactly how the types of relatedness measures determine participants' responses, regardless of how they are asked to make the judgment? To get some measure of the relative importance of each type of relatedness to the subjective judgment of relatedness, regardless of what is meant by "related"?  
      
    ~~Please also summarize the analysis of Hypothesis 3: what is most important in predicting cued recall of word pairs?~~ Hypothesis 3 is already summarized with 2. They’re the only one to complain about it, so pretty sure they just missed it.  
      
    Specific Comments:  
      
    -The acronyms are confusing. I had to make a reference note to keep track of which sort of relationship was intended by FSG vs. COS vs. LAS. It might make it easier to replace these acronyms with the actual names of the type of relatedness (e.g., replace "COS" with "thematic relatedness") and just state in the materials section which specific measure/database was used for each. CHANGED A FEW OF THESE FOR CLARITY, BUT I’M NOT CHANGING ALL OF THEM.

    -Given that JOLs are discussed in the introduction, it is confusing to refer to the examined judgments of relatedness as simply "judgments"; it leads the reader to think that JOLs were obtained in the study. Please clarify throughout (even using "JoRs" if desired).   
      
    -~~On page 22, the authors state "While previous research has shown that memory networks are divided into separate systems which handle storage and processing for meaning and association," Please provide some references~~. FIXED  
      
      
    Reviewer: 2  
      
    Comments to the Author  
    This research examined the relationships of norms for associative strength, semantic overlap, and thematic strength between word pairs with judgments of relatedness and cued recall performance involving those same pairs. Participants rated associative strength, semantic overlap, and thematic strength of word pairs in three separate blocks. When semantic overlap and thematic strength were low, norms for associative strength were found to strongly predict judgments of relatedness, whereas when thematic strength was higher, the predictiveness of associative strength was reduced. This suggests a competitive relationship between associative strength and thematic strength, with high levels of thematic strength associated with reduced influence of associative strength on participants’ judgments of relatedness. In contrast, when semantic overlap was high, the predictiveness of associative strength for relatedness judgments increased with increasing levels of thematic strength, suggesting a cooperative relationship between associative strength and thematic strength in this context. Similar results were obtained when examining the predictiveness of the three types of norms for cued recall performance. Finally, judgments of relatedness were found to be predictive of cued recall performance.   
      
    This research may well be of interest to researchers of semantic relatedness and cued recall performance, but I am not convinced it is of sufficient interest to the larger cognitive psychology community to warrant publication in Memory & Cognition. The authors do describe a theory of the organization of semantic memory networks that may be of interest to a broad readership, involving a three-tiered interconnected system of semantic, associative, and thematic relationships, but the connection between this theory and the specific data being presented is somewhat opaque (at least to me). If the authors can make a stronger case for how their data support this theory, this could greatly increase the impact of their work. Short of this, I would suggest that this paper may be better suited for a behavioral methods journal.   
      
    Apart from these global concerns, I have a few smaller issues, which I will address below, indexed by page number:  
      
    p. 11~~, Participants: The authors describe in general terms how participants were excluded, but specific criteria need to be stated. Some of this information is provided later as part of the data analysis, but if exclusion criteria are to be mentioned in the method section, they should be clearly specified.~~ Tried to clarify this  
      
    p. 12, ~~first full paragraph, first sentence: This sentence (i.e., “Stimuli were varied …”) is extremely confusing.~~ FIXED  
      
    p. 13~~, last paragraph: Participants are asked to rate how college students would respond to a given cue word, but it is not clear that all of the mTurk workers were college students or had ever attended college, so it is not clear that all of them have a sufficient basis for judgment on this question.~~ FIXED  
      
    p. 15, first paragraph: This level of detail regarding the organization of the datasheet is probably not necessary. NOT FIXING, I feel like this level of detail is important  
      
    p. ~~16, last paragraph, second sentence: Don’t need decimal points when relating numbers of participants.~~ FIXED  
      
    Reviewer: 3  
      
    Comments to the Author  
    Review of MS-ORIG-18-086  
    Investigating the Interaction between Associative, Semantic, and Thematic Database Norms  
    for Memory Judgments and Retrieval  
    Maxwell & Buchanan  
    Missouri State University  
      
    This ms is well-written and presents research on an interesting topic. My comments are listed below:  
    Major comments:  
    My main major comments has to do with Hypothesis 4. In order for JOLs to be predictive of recall, the information accessible to participants when making their JOLs must be diagnostic of recall. In considering the memory results, FSG was the strongest predictor of recall in both the associative and semantic conditions, with LSA being the strongest predictor of recall in the thematic condition. The prediction would follow that JOLs based on FSG in the associative and semantic conditions would then produce judgments slopes that were most predictive of recall. That finding was obtained in the association condition, but the FSG slope was not a significant predictor of recall in the semantic condition. Why not? Perhaps because COS was a close second predictor of recall in the semantic condition, JOLs that were based on COS in that condition produced the significant COS slope toward predicting recall. Perhaps more frequently basing JOLs on FSG, even in the semantic condition, produced inaccurate predictions that produced the nonsignificant FSG predictor of recall in that condition. Bottom line is, I would like to see some discussion about the available bases of the JOLs and how those are and are not diagnostic of the recall performance. One could suggest that FSG is more accessible to the average participant. Although the norming task used by Nelson et al produced sometimes-large sets of associates, they did tend to be clustered with high FSG for just 2 or 3 targets and much smaller associations spread over several other targets. Is FSG something that is more accessible to participants than COS or LSA? I have an intuitive sense that it is. I think the idea of what is accessible (ala Koriat) to participants when making their JOLs and whether that is consistent with the manipulation of the word pairs should be considered and discussed.   
      
    Minor comments:  
    ~~Line57/58 – Perhaps after word pairs, insert “such as the University of South Florida Word Association Norms” and then cite Nelson et al. Seems awkward to just site them for this statement as is.   
    Line 60: hyphenate context based (i.e., context-based)~~ FIXED  
    ~~Line 71 – can you site others associative word norms? You state several times that there are many of these, but only site Nelson’s~~. REFERENCED SWOW  
    ~~Line 130 – hyphenate text based (i.e., text-based) – also Line 134~~ FIXED  
    ~~Line 138 – awkward phrasing~~ FIXED  
    ~~Line 160 – change order to associative, semantic, and thematic systems~~ FIXED  
    ~~Line 201-202 – maybe expand on this idea of Koriat & Bjork (2005). Not immediately clear how their findings relate.~~ REPHRASED A BIT FOR CLARITY, BUT I’M NOT SURE HOW THIS DOESN’T MAKE SENSE  
    ~~Line 395/396 – awkward phrasing~~ FIXED  
      
      
      
    Reviewer: 4  
      
    Comments to the Author  
    Review for Memory and Cognition Manuscript: MC-ORIG-18-086  
      
    Summary:  
      
    The authors were interested in the relationship between word pairs associations and judgments of memorability and recall of the same word pairs. Participants studied 63 paired associates across three blocks. Each block comprised 21 word pairs and 21 corresponding judgments of learning (either via association framing, semantic framing, or thematic framing). Cued recall followed. Authors evaluated four hypotheses geared toward understanding if the three associative operational definitions reflect three independent networks of information, or if they share underlying cognitive processes.  
      
      
    Assessment:   
      
    Overall, I thought the paper addressed an interesting topic, was written well, and has impactful results in part because the authors used innovative experimental and statistical methods to evaluate the four primary hypotheses. Based on this analysis, I recommend the paper for publication. I have a few suggestions that I think the authors should consider before publication. These comments pertain more to big-picture issues that may be address with changes to the prose.  
      
    Ultimately, at its core, this paper provides an evaluation of two models of word knowledge—the three-tiered view and the dynamic attractor view. Often, however, this message gets lost in the mix. It might be helpful to the reader if the authors indicated exactly how experimental outcomes can help evaluate the theory. It isn’t clear, and if the authors do mention how data relate to theory, this mention is covered up by the deep historical overview in the intro or the myriad of analyses in the results. Thus, I can see this potential edit living in the intro, results section, or both.  
      
    I’m wondering if the hypotheses could be more specific in terms of theory—and if this approach might help connect data to theory. As it stands, the four hypotheses read as a-theoretical. Hypothesis 1 is pitched as a replication, hypotheses 2 and 3 are framed exclusively in terms of a statistic pattern (i.e., predicted interactions), and hypothesis 4 is pitched as a regression analysis. As I read through the results, I felt there was a disconnect between the detail and rigor of the statistical analyses (for which I commend the authors) and how each analysis contributed specifically to a deeper understanding of theory. For example, perhaps the mere existence of interactions, regardless of how the interactions work, indicates support for the dynamic model and evidence against a tiered model. By contrast, maybe the exact nature of the interaction is of theoretical interest. It isn’t clear.  
      
    Considering the above critique, I struggled to see the connection between the three-tiered model as it pertains to reading processing. THREE TIERED MODEL FITS IN AS PART OF SEMANTIC BUBBLE, BASICALLY SEMANTIC BUBBLE HAS THREE PARTS – ADDED SOMETHING ABOUT THIS TO DISCUSSION. I wasn’t sure if the authors were testing the triangle model of semantics, orthography, and phonology, or if they were merely using the triangle model as an example of how association in memory might work (with association, semantic, and thematic as the stand-in factors. My guess is the latter. If so, why even mention the details of the original triangle model? It feels a little tangential. Changed some sentences in the discussion to make this flow a bit better, removed some of the details.  
      
    Other Comments:  
      
    I next list some minor points and suggestions.   
      
    ~~On page 13, it is unclear to me what the three phases are in the Procedure section (the authors also call them “sections”). It might be helpful to always refer to each phase as a “phase” rather than a section, considering that earlier the authors referred to there being “three phases”. My guess is the three phases are (1) the study/judgment phase, (2) the distractor phase, and (3) the test phase. But maybe not? Perhaps make this more explicit for the reader~~. FIXED  
      
    ~~In the paragraph at the bottom of page 13, the authors include description of the judgment task for the associative and semantic conditions but not for the thematic condition. Was there any reasoning for omitting the thematic judgment description? I was able to see what was asked of subjects in the thematic condition by visiting the URL listed in the manuscript, but the thematic judgment should be described in the text~~. FIXED  
      
    ~~Page 7: There’s something off in the statement that starts with “Potentially,…”. Subject – verb agreement?~~ FIXED  
      
    ~~Page 8: There are two instances of the word “represent” when only one is needed.~~ FIXED  
      
~~Page 8: The authors refer to cued-recall as a matching task (i.e., “…match the target with the cue on a recall task”). I’m not sure this is a correct characterization. Matching evokes some form of recognition memory. From what I understand, memory was tested using cued recall.~~ FIXED  
      
    ~~Page 10: “…levels other…” should be “…levels of other…”~~ FIXED  
      
    ~~Page 12: The sentence that starts with “Stimuli were varied…” is a little confusing; perhaps because of the overuse of the term “varied” and “variable” in the same sentence.~~ FIXED  
      
    ~~Page 13: “…with given….” should perhaps be “…with a given…” ?~~ FIXED  
      
    ~~Page 22: I’m unclear about the phrase “place world”. Please revise this.~~ FIXED