

Multiple Short Tests or One Long Test: Which Is Better for Learning?

Shaun Boustani, Hilary Don, David Shanks

Division of Psychology and Language Sciences, University College London



Background

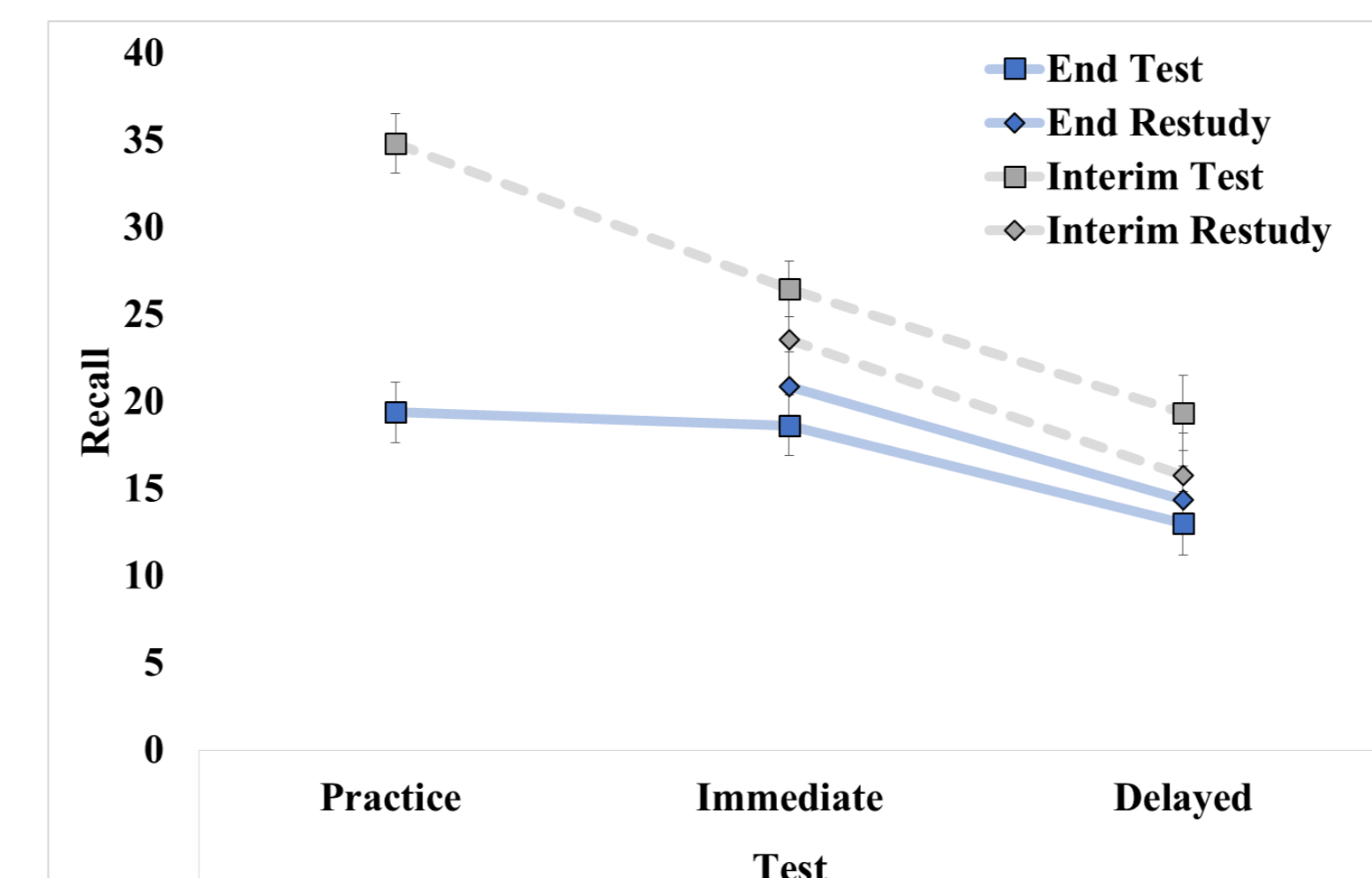
- Testing through retrieval practice can enhance memory and future learning.
- The grain size of recall hypothesis suggests that multiple short tests throughout learning will result in higher practice test recall and better long term retention than a single long test at the end.
- Research has so far found a Grain Size Effect (GSE) on practice tests but not cumulative assessment (Wissman & Rawson, 2015).

The Current Study

- Research on the GSE has used complex materials (i.e., prose, trivia, lectures) and has not assessed whether the potential benefit of smaller grains is specific to tests or also occurs for restudy. Additionally, few studies have assessed the GSE in the confirmed presence of a Forward Test Effect (FTE).
- We evaluated the grain size hypothesis using lists of related and unrelated words and exposure-matched restudy and FTE controls groups.
- Primarily we wanted to explore whether a GSE existed using word lists, and if so, ascertain whether it is:
 - Specific to recall
 - Dependent on material relatedness
 - Persisted across a delay

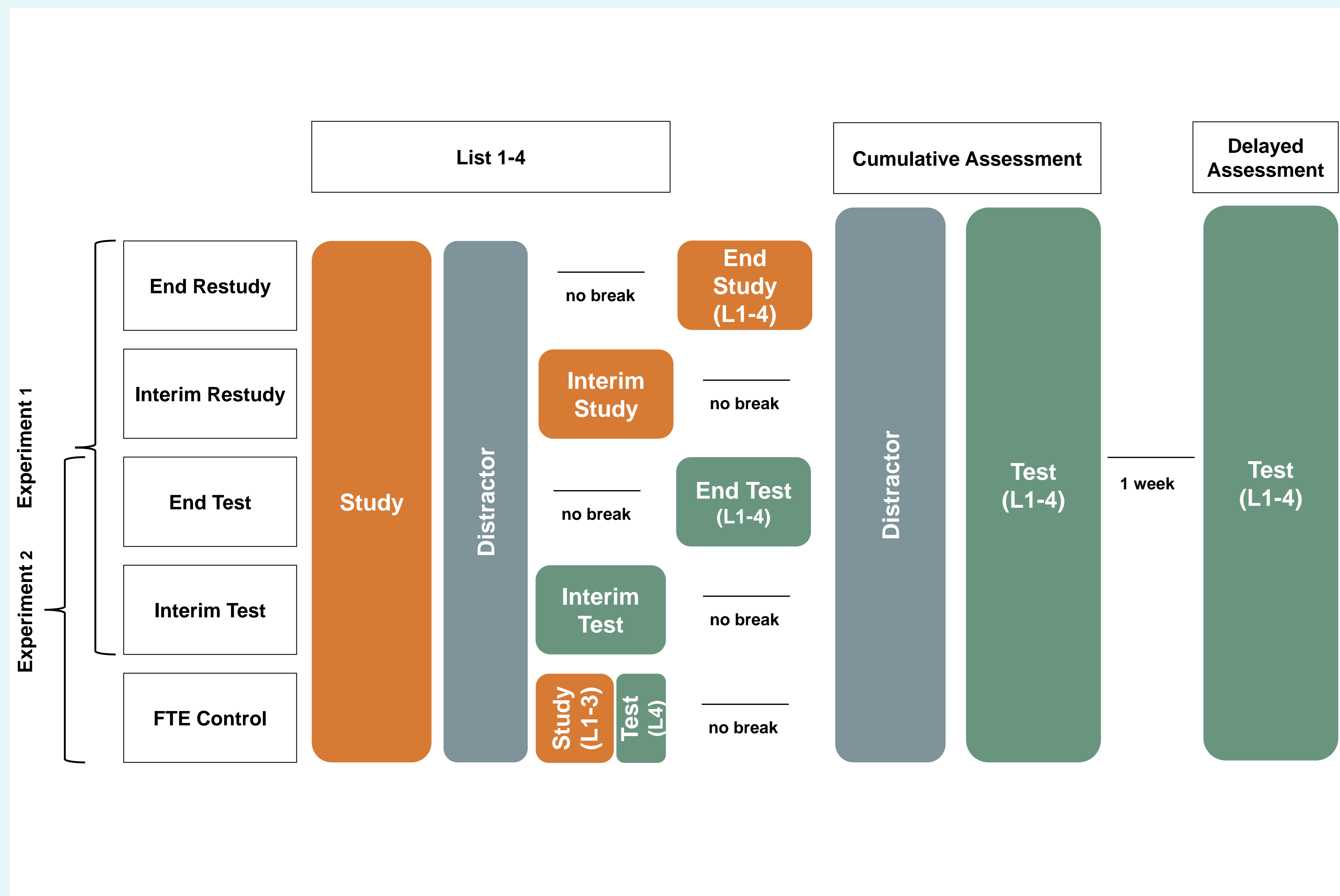
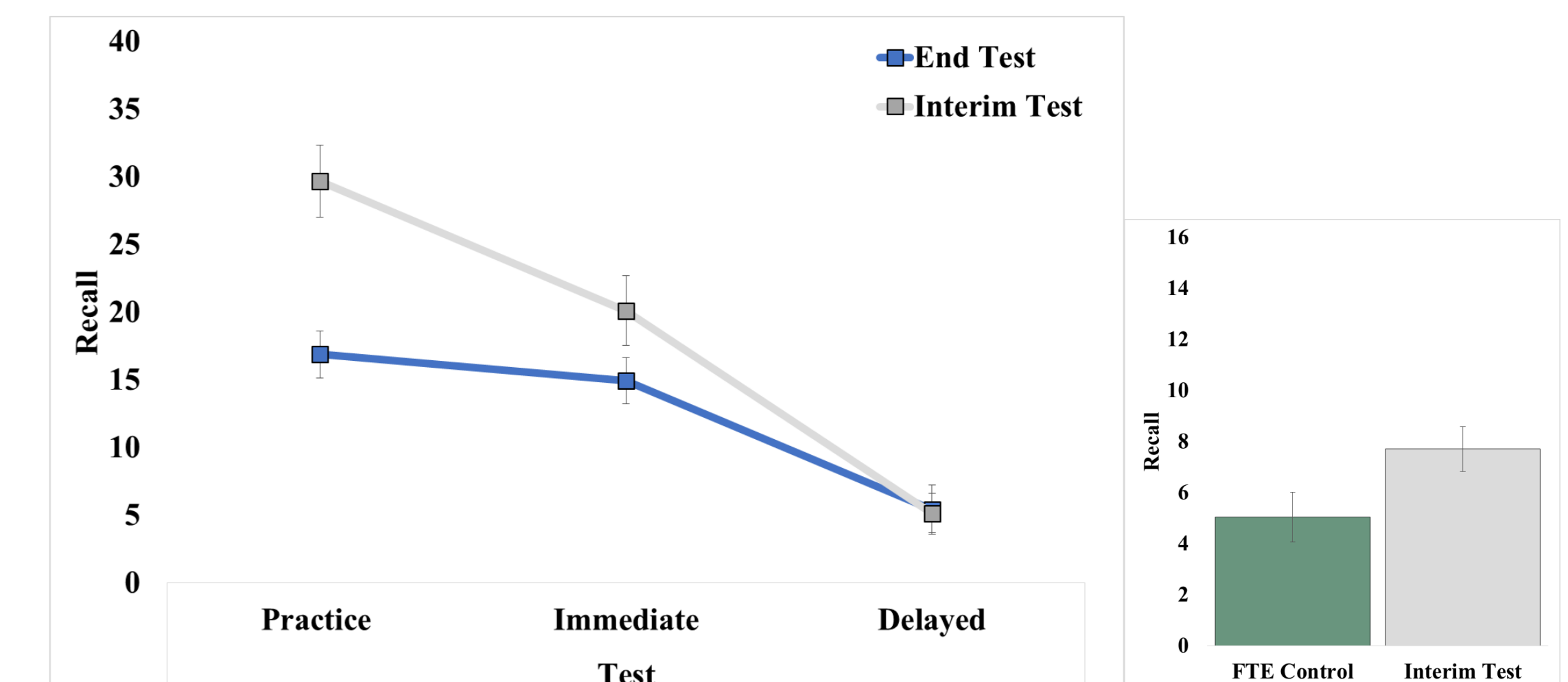
Experiment 1 (N = 346)

- Related words
- Interval Task (Test; Restudy) vs Task Schedule (Interim; End).



Experiment 2 (N = 149)

- Unrelated words
- Interim Test, End Test, Control group assessing the presence of an FTE.



Key Findings

- For related and unrelated materials, we found a **large GSE for practice tests** (Exp. 1: $p < .001$, $\eta^2_p = .457$; Exp. 2: $p < .001$, $\eta^2_p = .275$)
- Experiment 1 found an interaction between interval task and schedule, with a **larger effect of grain size on tests than restudy** on both immediate ($p = .01$, $\eta^2_p = .018$) and delayed assessments, ($p = .023$, $\eta^2_p = .023$)
- Experiment 2 found a **smaller but still significant GSE** on the immediate assessment, ($p = .034$, $\eta^2_p = .045$) but **no GSE on the delayed assessment**, ($p = .579$, $\eta^2_p = .008$).
 - Experiment 2 also found an **FTE**, with superior criterial list recall in the Interim Test than the Control group ($p < .001$, $\eta^2_p = .141$)

Conclusion

- Multiple short tests are better for learning than one long test.
- However, the effect appears to degrade over longer retention intervals. There is surprisingly substantial forgetting from interim practice tests to final cumulative recall.

Reference: Wissman, K. T., & Rawson, K. A. (2015). Grain size of recall practice for lengthy text material: Fragile and mysterious effects on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 439–455. doi.org/10.1037/xlm0000047