Does Test Difficulty Matter?

Examining the Influence of Stimulus Difficulty on Cues that Affect Learning Confidence in Young Adults

Yan Lin Lee & Tyler Small | Faculty Mentor: Jarrod Hines
Department of Psychology

Introduction

When you have many chances to study information, learning rates are maximized by focusing on currently unlearned items.[1,2] To do this, one must monitor the success of prior study attempts accurately. For example, when studying for a test, feeling less confident on certain topics relative to others will alert individuals of the need to focus more on the less familiar topics.

This *feeling of confidence* is called a **judgment of learning (JOL)** and is defined as *the confidence one has in the ability to remember studied material*. Metacognitive judgments like JOL reflect perceptions of performance and can potentially be influenced by multiple sources of information about one's current state of learning.[3,4] Such sources are called metacognitive cues.

Prior JOL research has emphasized the powerful role of one cue, memory for past test (MPT), on subsequent JOLs.[5,6] MPT reflects one's remembrance of prior memory success or failure during a test. Newer research supports the roles of other metacognitive cues on JOLs in addition to that of MPT,[4] such as: (1) assessments of item difficulty made after an initial study opportunity, (2) subjective memory response confidence during the first test, (3) objective response times during the first test, and (4) subjective response time estimates (how fast they think they responded to test questions).

In the current research, we investigated the relative impact of the above-listed cues on subsequent JOL formation as a function of stimulus difficulty. Although difficulty level is widely-used as a manipulation in cognitive research, this is not the case for studies examining JOL formation. Most prior JOL research used word pairs that were homogenous in their difficulty level. We hypothesized that the relative impact of metacognitive cues on JOL formation would differ as a function of stimulus difficulty level (easy vs. hard).

Methods

Participants

Participants were UWEC students (M=2, W=22; age *M*=19.83 years, *SD*=1.79) who signed up via the SONA online research pool in exchange for course credit or extra credit.

Procedure

Participants completed three study-test phases. In each phase, they studied word pairs and were then tested on their memory for each pair. We manipulated the difficulty level of the word pairs (easy or hard). All participants were exposed to the same set of word pairs, which consisted of 30 easy-to-remember (e.g., REASON-THINK) and 30 difficult-to-remember (e.g., CLEMENCY-IDIOM) word pairs.

Phase 1: Study

- 60 word pairs of unrelated nouns were presented in a randomized order. Study time for each word pair was self-terminated.
- After studying each item, participants made a JOL on a scale of 0-100% memory confidence.

Phase 1: Test

- 120 word-pairs (60 intact, 60 rearranged) were presented in a randomized order.
- Using designated keys on a keyboard, they entered "YES" for a recognized pair and "NO" for an unrecognized pair.
- After each test trial, participants estimated (1) a confidence judgment on a scale of 0-100% (i.e., how confident they were in the accuracy of their prior recognition response), and (2) a response time estimate (on a scale of 0-10s, for the prior recognition response).

Phases 2 and 3: Study and Test

- The same stimuli and procedure as in Phase 1 were used.

Note: Due to ceiling-level performance in Phase 3, those data were not included in the current analyses.

Results

Table 1: Means Values for Measured Objective and Metacognitive Cues by Task Phase and Stimulus Difficulty

Cue	Stimulus Difficulty	Phase 1 *M (SD)*	Phase 2 *M (SD)*
Study Time (ST)	Easy	5.24 (5.52)	1.35 (0.69)
	Hard	7.14 (6.45)	3.71 (2.35)
Judgment of Learning (JOL)	Easy	65.37 (22.11)	89.49 (11.54)
	Hard	31.23(18.49)	51.56 (20.12)
Confidence Judgment (CJ)**	Easy	92.23 (6.48)	95.73 (4.62)
	Hard**	60.07 (15.26)	71.12 (15.35)
Accuracy (ACC)	Easy	94.10 (8.20)	96.88 (8.20)
	Hard	67.78 (8.20)	76.53 (8.20)
Response Time (RT)	Easy	1.91 (0.46)	1.35 (0.30)
	Hard	3.05 (1.09)	2.43 (0.95)
Response Time Estimate (EST)*	Easy	1.82 (1.24)	1.82 (1.32)
	Hard	2.33 (1.56)	2.29 (1.59)
Difference between EST and RT (DIFF)	Easy	-0.14 (1.25)	0.41 (1.27)
	Hard	-0.76 (1.69)	-0.17 (1.58)

Note 1. For all variables but EST, there were main effects of both Phase and Stimulus Difficulty. *Main effect of Stimulus Difficulty. **Phase x Stimulus Difficulty interaction. *Note 2.* JOL, CJ, and ACC computed in percent; ST, RT, EST, and DIFF computed in seconds.

Table 2: Multi-level Regression Predicting Phase 2 Judgment of Learning (JOL)

Effect	Estimate	*SE*	*df*	*t*
Intercept	**21.6491**	**3.3798**	**15.7**	**6.41*****
Difficulty	**33.9702**	**1.0276**	**1076**	**33.06*****
Easy items were given about 34% higher JOLs.				
JOL (Within-Ss)	**0.3456**	**0.09426**	**1100**	**3.67*****
For every 1% increase in item-level P1 JOL, there was a corresponding 0.35% increase in item-level P2 JOL.				
ACC (Between-Ss)	**124.79**	**46.7891**	**120**	**2.67*****
On average, people who were more confident in P1 were more confident in P2.				
CJ (Within-Ss)	**0.282**	**0.0945**	**262**	**2.98*****
For every 1% increase in item-level P1 CJ, there was a corresponding 0.28% increase in item-level P2 JOL.				
CJ (Between-Ss)	**1.0923**	**0.3386**	**502**	**3.23*****
For every 1% increase in a person's P1 mean CJ, there was a corresponding 1.1% increase in a person's P2 mean JOL.				
DIFF x CJ (Between-Ss)	**-0.8224**	**0.3599**	**243**	**-2.29****
Estimate for easy items = 0.71; estimate for hard items = 0.72. Means differed statistically, but this holds little practical significance.				
RT (Between-Ss)	**-9.35**	**5.719**	**227**	**-1.64**
For every 1s increase in a person's P1 mean RT, there was a corresponding 9% decrease in a person's mean P2 JOL. When broken down by difficulty, the effect was significant** for easy items, but not for hard items.				
ST (Within-Ss)	**-1.16**	**.663**	**1096**	**-1.74***
For every 1s increase in item-level P1 ST, there was a corresponding 1.2% decrease in item-level P2 JOL. When broken down by difficulty, the effect was significant*** for both easy and hard items.				

Note. *p <.10, **p < .05, ***p<.01.

As shown in Table 2, regression analyses revealed no support for our hypothesis. Stimulus difficulty did not influence the relative impact of cues on JOL formation.

Discussion

We designed this study to investigate the relative impact of metacognitive cues on JOL formation as a function of stimulus difficulty. We did not find the relationships between our measured cues and subsequent JOL to differ due to this factor, based on a lack of higher-order interactions, and this leaves our hypothesis unsupported.

Nevertheless, our results showed consistent main effects of Phase and Stimulus Difficulty. Participants spent less time studying and responded faster during tests in Phase 2 than in Phase 1; reported higher JOLs and CJs in Phase 2 than in Phase 1; and achieved higher test accuracy in Phase 2 than in Phase 1.

Participants spent more time studying hard word pairs and responded slower during tests for hard word pairs than they did for easy word pairs; reported lower JOLs and CJs for hard word pairs than for easy word pairs; achieved lower test accuracy for hard word pairs than for easy word pairs; and estimated longer response time during tests for hard word pairs than for easy word pairs.

We are left to explore alternative explanations for the lack of influence of stimulus difficulty. One possibility is that, although our participants received explicit instructions to make metacognitive judgments (e.g., CJs), these cues were not utilized consciously when making JOLs. The formation of JOLs may be relatively stable across stimulus difficulty due to the presumably large amount of prior practice subjects had in estimating their confidence prior to our study. It is possible, perhaps even likely, that such estimates are so well-practiced that they have become essentially automatic and somewhat unconscious, and thus are relatively insensitive to variables like stimulus difficulty. For example, although our manipulation of stimulus difficulty did influence memory confidence and accuracy, we did not change the way in which that information was used to generate JOLs.

Future Directions

Our next study will involve a modified version of the current word pair recognition task that makes use of (1) more word pairs of both difficulty levels, and (2) a third level of intermediate difficulty. Based on the current findings, we predict that having more items will reduce potential ceiling effects in individual responses and hence enhance the variability in our data. This will allow for a more precise examination of the roles of metacognitive cues on JOL formation.

References

1. Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General, 122,* 47-60. doi:10.1037/0096-3445.122.1.47
2. Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676-686. doi:10.1037/0278-7393.14.4.676
3. Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349-370. doi:10.1037/0096-3445.126.4.349
4. Hertzog, C., Hines, J. C., & Touron, D. R. (2013). Judgments of learning are influenced by multiple cues in addition to memory for past test accuracy. *Archives of Scientific Psychology, 1,* 23-32. doi:10.1037/0882
5. Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 238-244. doi:10.1037/0278-7393.33.238
6. Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58,* 19-34. doi:10.1016/j.jml.2007.03.006

Acknowledgements

This research is supported by the Office of Research and Sponsored Programs at UWEC.