



Can very small font size enhance memory?

Vered Halamish¹

Published online: 3 May 2018
© Psychonomic Society, Inc. 2018

Abstract

Presenting information in a perceptually degraded format sometimes enhances learning outcomes. However, earlier studies in which words were presented in large or small fonts in a paradigm that also involved item-by-item judgments of learning (JOLs) consistently yielded no mnemonic benefit of small fonts. Can small font size enhance memory under hitherto unexamined conditions? A series of 11 experiments was conducted to examine systematically the effect of font size on memory for words and whether it depends on the strength of the font size manipulation, whether JOLs are solicited, the format of the test, and study time. The resulting data were meta-analyzed. Results yielded a u-shape relationship between font size and memory. Compared to intermediate fonts, there was a memory advantage for words presented in large fonts but also for words presented in very small fonts. However, the requirement to provide a JOL moderated the benefit of the very small font, which was eliminated when JOLs were solicited. Test format and study time did not moderate the effect of font size on memory. JOLs were insensitive to the u-shape relationship and consistently increased with font size. These findings provide support for the notion that perceptually degraded materials can enhance learning outcomes, but also highlight the importance of systematic investigation of moderators. The results shed new light on the inconsistent effects of manipulations of perceptual degradation on learning outcomes observed in earlier studies.

Keywords Memory · Font size · Perceptually degradation · Desirable difficulties · Judgments of learning

Introduction

Do we remember learning materials better when they are presented in a format that is more difficult to read or when they are presented in a format that is more easy to read? Memory for study materials is clearly affected by difficulty of the content, learning strategy, and cognitive ability, but is it also affected by extraneous factors such as font size or type? Currently, there is no clear answer to these questions, because the evidence is mixed. One line of research that focused on the font size of to-be-remembered words consistently suggested that font size does not affect recall (e.g., Rhodes & Castel, 2008), but a recent meta-analysis that was based on these studies suggested that there is nevertheless a subtle memory advantage for the larger font words (Luna, Martin-Luengo, &

Albuquerque, *in press*). In contrast, a second line of research that focused on other perceptual features of learning materials such as font type or clarity suggested that, in some cases, presenting materials in a perceptually degraded format can enhance rather than impair learning (e.g., Diemand-Yauman, Oppenheimer, & Vaughan, 2011). Can small font size similarly enhance memory? The current research investigated whether, under hitherto unexamined conditions, presenting words in a small font enhances memory.

Font size and memory for words

In a series of experiments, Rhodes and Castel (2008) presented words on a computer screen in 18-point or 48-point font. Participants were asked to memorize the words and to provide a judgment of learning (JOLs) for each studied word by estimating how confident they were that they would later be able to recall that word. They expected font size to have relatively little impact on free recall because, they argued, memory is predominately influenced by processing of the meaning of the stimuli (e.g., Craik & Lockhart, 1972). Their main interest was in whether learners' JOLs would reflect that. Rhodes and Castel found that larger words were predicted to be

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13421-018-0816-6>) contains supplementary material, which is available to authorized users.

✉ Vered Halamish
vered.halamish@biu.ac.il

¹ School of Education, Bar-Ilan University, Ramat-Gan, Israel

remembered better than smaller ones, but that this was a metacognitive illusion because, as expected, recall was not affected by font size.

Over the last decade, these findings have been replicated in several similar studies (Hu, Li, Zheng, Si, Liu, & Luo, 2015; Kornell, Rhodes, Castel, & Tauber, 2011; Luna, et al., *in press* McDonough & Gallo, 2012; Miele, Finn, & Molden, 2011; Mueller, Dunlosky, Tauber, & Rhodes, 2014; Susser, Mulligen, & Besken, 2013). These studies used the same basic paradigm as the original study, with only minor variations in materials and test format (e.g., word pairs and a cued-recall test in Rhodes & Castel, 2008, Experiment 3), study time (between 2 and 5 s), procedure (delayed JOLs, Luna et al., *in press*) and font sizes (125-point vs. 25-point, McDonough & Gallo, 2012; 70-point vs. 9-point Chinese characters, Hu et al., 2015; various font sizes depending on the participants' personal screen settings, Kornell et al., 2011). While no individual study showed a significant difference, a recent meta-analysis by Luna et al. (*in press*) revealed a small memory advantage for items presented in large font over small font. This advantage was much smaller than the one predicted by learners, reflecting a mismatch between the effect of font size on memory and metamemory.

Perceptual degradation and learning

There is, however, evidence that presentation of learning materials in a perceptually degraded format can sometimes improve learning. This evidence comes mainly from studies that used perceptual manipulation other than of font size.

One line of evidence comes from studies of perceptual interference in which a brief (i.e., 100-ms) presentation of a word is followed by a presentation of a pattern, which causes backward masking. Perceptual interference was found to enhance subsequent recognition and recall relative to longer presentations (i.e., 2.5 s) without interference (Besken & Mulligan, 2013; Hirshman & Mulligan, 1991; Hirshman, Trembath, & Mulligan, 1994; Mulligan, 1996).

More recent evidence comes from studies that used complex, educationally relevant learning materials. In a noteworthy study by Diemand-Yauman et al. (2011), university and high-school students studied textual materials that were presented in an easy-to-read font (standard black font) or in a difficult-to-read font (small, non-standard, gray font). Performance on subsequent tests suggested that studying the materials in a difficult-to-read font produced better learning outcomes than studying in an easy-to-read font. This finding was replicated in a number of subsequent studies (Eitel, Kühl, Scheiter, & Gerjets, 2014, Experiment 1; French et al. 2013; Lehmann, Goussios, & Seufert, 2016; Seufert, Wagner, & Westphal, 2017; Weissgerber & Reinhard, 2017; Weltman & Eakin, 2014).

These findings are consistent with the idea that difficulties can be desirable for learning (Bjork, 1994; Bjork & Bjork, 2011). Often, encoding manipulations that make learning slower and more difficult actually enhance long-term retention and transfer of learning. These manipulations include, for example, spacing, variation, and interleaving, and have been termed “desirable difficulties” (Bjork, 1994). The recent evidence suggests that presenting information in a perceptually degraded format can also be a desirable difficulty.

However, other studies failed to replicate the beneficial effect of difficult fonts, finding either no effect (e.g., Eitel et al., 2014, Experiments 2–4; Eitel, & Kühl, 2016; Pieger, Mengelkamp, & Bannert, 2016; Rummer, Schweppe, & Schwede, 2016; Strukelj, Scheiter, Nyström, & Holmqvist, 2016; Yue, Castel, & Bjork, 2013), or the opposite effect, impaired learning when studying in difficult fonts (e.g., Lonsdale, Dyson, & Reynolds, 2006; Miele & Molden, 2010; Yue et al., 2013). Nevertheless, the findings of Diemand-Yauman et al. (2011) and others suggested that presenting materials in a perceptually degraded format can, under some conditions, serve as a desirable difficulty.

More recent studies have therefore investigated the conditions under which perceptually degraded materials enhance learning (Dunlosky & Mueller, 2016; Oppenheim & Alter, 2014; Kühl, Eitel, Scheiter, & Gerjets, 2014; Weissgerber & Reinhard, 2017). For example, Lehmann et al. (2016) observed that perceptually degraded fonts improved learning only for learners with a high working memory capacity; Weissgerber and Reinhard (2017) observed that perceptually degraded fonts enhanced long-term, but not short-term memory; Halin, Marsh, Hellman, Hellstrom, and Sorqvist (2014) observed that perceptually degraded fonts improved performance only when there was distracting background noise; and Katzir, Hershko, and Halamish (2013) observed that smaller-than-standard fonts enhanced fifth graders' reading comprehension, but impaired that of second graders.

Of most relevance to the current research are two other studies that found, using a procedure broadly similar to that of Rhodes and Castel (2008), that perceptually degraded presentation can be a desirable difficulty for the learning of single words. Sungkhasettee, Friedman, and Castel (2011) presented single words upright or upside down. Participants predicted equivalent memory for upright and inverted words, but free recall was higher for the inverted words, suggesting that inverted presentation is a desirable difficulty. Rosner, Davis, and Milliken (2015) examined the effect on memory of perceptual blurring of words, and observed better recognition memory for blurred words than for clear words, although an earlier study (Yue et al., 2013) failed to obtain a benefit from blurring. This study is discussed in more detail below.

Why does presenting learning materials in perceptually degraded format enhance memory? A common suggestion is that perceptually degraded formats function as a metacognitive cue to

allocate more cognitive resources or to enhance cognitive engagement in the learning task. As a result, processing of degraded material is more effortful or deeper and this enhances memory (Alter, Oppenheimer, Epley, & Eyre, 2007; Diemand-Yauman et al., 2011; Hirshman & Mulligan, 1991; Mulligan, 1996). This explanation is rather general and suggests that different perceptual manipulations should have similar effects on learning outcomes. In contrast, Weissgerber and Reinhard (2017) proposed, based on the concept of transfer-appropriate-processing (McDaniel & Butler, 2011), that different perceptual manipulations might invoke different processes during encoding and that their effect on subsequent memory would depend on the match between the processes invoked and the demands of the memory test. Clearly, more research is needed to fully understand the mechanisms underlying the potential effects of perceptually degraded presentation formats on memory.

Font size and memory for words: potential moderators

The literature reviewed above provides ample evidence that presenting materials in perceptually degraded formats can enhance memory and learning outcomes and act as a desirable difficulty, but also provides clear evidence that this effect is far from robust. Considering this evidence, the lack of evidence for a beneficial effect of small font size is surprising. Why does presenting words in a small font not enhance memory in the same way as inverted (Sungkhasettee et al., 2011) or blurred (Rosner et al., 2015) presentations?

The current research examined the possibility that small font size can enhance memory, but that the conditions under which it does so had simply not yet been examined. It focused on four potential moderators of the effect of font size on memory: strength of the font size manipulation, whether JOLs are solicited or not, the test format, and study time.

The strength of the font size manipulation The small fonts used in earlier studies may have been ineffective simply because they were not sufficiently small to trigger the processes that led to enhancement of memory. Rhodes and Castel (2008) and most follow-up studies used 48-point and 18-point fonts as the large and small font sizes, respectively. The current research was driven by the conjecture that whereas 48-point font can indeed be described as relatively large and easy to process, 18-point font is best described as a standard or medium size, rather than small, and is not difficult to process. Critically, 18-point font might not be small enough to induce the cognitive engagement and effortful processing that render other forms of perceptual degradation (e.g., word inversion or unusual font type) desirable difficulties. Consistent with this argument, Rosner et al. (2015) observed that the level of blurring moderated its effect on recognition memory. A benefit of blurred (over clear) fonts was obtained only for relatively

higher levels of blurring. Similarly, using textual materials, Seufert et al. (2017, Experiment 2) demonstrated that increasing the level of font difficulty, up to the point where the text became illegible, enhanced recall and transfer performance.

Solicitation of JOLs In the study by Rhodes and Castel (2008) and follow-up studies, participants were asked to predict the chance that they would remember the words they were studying (i.e., provide a JOL) during the learning phase, on an item-by-item basis. However, recent studies suggest that soliciting memory judgments during learning alters the encoding processes that would otherwise occur (e.g., Mitchum, Kelley, & Fox, 2016; Nguyen & McDaniel, 2016; Schmidt & Schmidt, 2017; Schnaubert & Bodemer, 2017; Soderstrom, Clark, Halamish, & Bjork, 2015; Witherby & Tauber, 2017; Zechmeister & Shaughnessy, 1980; for similar effects of judgments made during the test see Double & Birney, 2017; Naveh-Benjamin & Kilb, 2012). These studies usually suggested that JOLs might improve memory (e.g., Soderstrom et al., 2015; Witherby & Tauber, 2017), although several other studies reported no effect (Benjamin, Bjork, & Schwartz, 1998; Tauber & Rhodes, 2012) or a detrimental effect (Mitchum et al., 2016) of JOLs. Moreover, some studies suggested that JOLs eliminate differences between conditions that are observed when JOLs are not solicited (Begg Vinski, Frankovich, & Holgate, 1991; Besken & Mulligan, 2013; Matvey, Dunlosky, & Guttentag, 2001; Rosner et al., 2015; Soderstrom et al., 2015). For example, Besken and Mulligan (2013) demonstrated that the mnemonic benefit of perceptual interference was eliminated when item-by-item JOLs were solicited, and Rosner et al. (2015) similarly demonstrated that the benefit of high levels of blurring for recognition memory was eliminated when JOLs were solicited. It is possible that the processes involved in making item-by-item JOLs encourage additional processing, which overlaps with the processes that are responsible for mnemonic benefits of perceptual interference or blurring. As Rosner et al. (2015, p. 20) concluded, “desirable difficulty effects in remembering may be difficult to observe when item-by-item JOLs are made at the time of encoding.” This evidence is consistent with the possibility that the solicitation of JOLs in previous studies hindered a potential effect of font size on memory.

Test format Rhodes and Castel (2008), and most follow-up studies, examined the effect of font size on cued or free recall performance. It is possible that recall tests are not sensitive to the beneficial effects of small font size. Recognition tests might be more sensitive to such effects, if they exist. Nairne (1988) suggested that perceptual manipulations affect the processing of surface-level aspects of the word that aids subsequent recognition, but not recall, which relies more on item elaboration. Indeed, perceptual degradation and interference manipulations were demonstrated to have benefits for

recognition more than for recall (Hirshman & Mulligan, 1991; Mulligan, 1996; Nairne, 1988; Rosner et al., 2015). McDonough and Gallo (2012) did examine the effect of font size on recognition and found no effect, consistent with the results for recall. However, they used relatively large font sizes and solicited item-by-item JOLs, which might have hindered the effect of font size on recognition memory.

Study time Earlier studies that examined the effect of perceptual degradation of words (small font or blurring) on memory varied in study time, which ranged from 0.5 to 5 s per word (between studies). The study in which a benefit of perceptual degradation was observed (Rosner et al., 2015) presented words for 1 s each. Perceptual interference effects were also observed with very brief presentation times (e.g., Mulligan, 1996). These findings suggest that the benefits of perceptual degradation may only be apparent when the study time is relatively short. It might be that relatively elaborative processing takes place spontaneously with longer but not with shorter study time, unless information is presented in a perceptually degraded format.

The current research

The current research was designed to investigate the effect of font size on memory for words and whether it depends on the strength of the font size manipulation, whether JOLs are solicited, the format of the test, and study time.

First, a series of 11 experiments were conducted to meet this goal. Table 1 presents an overview of these experiments. In all experiments, participants studied single words that were presented one-by-one on a computer, and were later tested on these words. The effect of the strength of the font size manipulation was examined by using three font sizes within-participants in all eleven experiments, the 48-point and 18-point sizes used in the earlier studies (e.g., Rhodes & Castel, 2008), and 5-point, a much smaller size. Pretesting suggested that the 5-point font was the smallest generally legible font size. Hereafter, the 5-point font is referred to as a very small font size. Solicitation of JOLs, test format (free recall vs. recognition), and study time (5 s vs. 0.5 s per word) were systematically manipulated across Experiments 1–8. To preview, in only two of these experiments was an effect of font size observed. Experiments 9 and 10 were attempts to replicate these effects. Experiment 11 directly examined the moderating role of solicitation of JOLs by manipulating this factor within the experiment, in addition to font size.

Next, a set of small-scale meta-analyses that included data from all 11 experiments are reported. The purpose of the meta-analyses was two-fold: first, to examine whether there are any font size effects that emerge when experiments are combined even though they might not be observed, or might not be consistently observed, in

single experiments (following Luna et al., *in press*); and secondly, to assess the potential moderating role of solicitation of JOLs, study time, and test format, that were usually manipulated between experiments (with the exception of the direct manipulation of JOLs solicitation in Experiment 11).

Experiments 1–4

Experiments 1–4 examined the effect of font size on memory using three different font sizes, when JOLs were solicited during study. Study time was relatively long (5 s) in Experiments 1–2 and relatively short (0.5 s) in Experiments 3–4. The test was a free-recall test in Experiments 1 and 3, and a recognition test in Experiments 2 and 4. The retention interval was modified according to the study time and test format, to avoid ceiling or floor effects.

Experiment 1: Free recall, long study time, JOLs

Method

Participants In Experiments 1–10, there was no a-priori estimate of effect size. Using rule of thumb, the sample size for Experiments 1–3 was set at about 30, and for Experiments 4–10 this was increased to about 42. Thirty-two students (19 women, age range: 18–32 years, mean age = 24.07 years) from the University of Haifa participated in Experiment 1. Participants were tested individually and received monetary compensation for their participation.

Materials Materials consisted of 48 nouns taken from Drori and Henik's (2005) norms for Hebrew words. The words were randomly divided into three sets of 14 items, matched for mean estimated familiarity (1–7 scale; $M = 3.91$, $SE = .07$), mean estimated concreteness (1–7 scale; $M = 6.02$, $SE = .10$), mean number of letters ($M = 4.35$, $SE = .18$), and mean number of syllables ($M = 2.23$, $SE = .08$). The remaining six items served as primacy and recency buffers, and were excluded from all analyses. Materials (in Hebrew) are available from the author upon request.

Procedure Participants were seated at a fixed distance of 23 in. from a 19-in. screen using a chin rest. They were asked to study words for a later memory test and were informed that the words would be presented in various font sizes. The words were then presented one at a time on the computer screen in black Arial font on a white background for 5 s each. Words from the three sets were intermixed and presented in a random order, with the restriction that no more than two items from the same set were presented consecutively. Words from one set were presented in 48-point font, words from another set were presented in 18-

Table 1 Overview of the methods, descriptive statistics, and summary of main results of Experiments 1–11

Experiment	Test format	Study time per word (sec)	Retention interval	JOLs solicited?	Mean proportion recalled/recognized* (SD)			Mean JOL (SD)			Effect of font size**	
					48-pt	18-pt	5-pt	48-pt	18-pt	5-pt	On memory	On JOLs
1	free recall	5	5 min	Yes	20.76 (11.24)	18.97 (11.98)	21.88 (12.30)	57.42 (20.91)	49.78 (20.95)	41.14 (23.94)	No	Yes 48>18>5
2	recognition	5	2–5 hrs	Yes	87.62 (9.37)	85.24 (12.71)	83.33 (11.46)	47.78 (23.75)	41.74 (20.32)	33.55 (19.32)	No	Yes 48,18>5
3	free recall	0.5	10 s	Yes	23.33 (13.38)	22.14 (14.08)	25.71 (12.38)	51.83 (14.88)	45.45 (14.37)	39.47 (13.79)	No	Yes 48>18>5
4	recognition	0.5	10 s	Yes	83.33 (15.67)	81.97 (15.95)	78.51 (17.56)	57.68 (24.86)	48.06 (22.07)	38.11 (21.71)	No	Yes 48>18>5
5	free recall	5	5 min	No	21.97 (14.13)	18.71 (13.47)	22.45 (12.48)	-	-	-	No	n/a
6	recognition	5	2–5 hrs	No	74.73 (18.01)	76.74 (16.45)	78.57 (14.09)	-	-	-	No	n/a
7	free recall	0.5	10 s	No	21.09 (11.37)	15.31 (11.11)	19.05 (11.98)	-	-	-	Yes 48>18	n/a
8	recognition	0.5	10 s	No	67.41 (14.02)	66.77 (18.87)	74.08 (16.13)	-	-	-	Yes 5>18	n/a
9	free recall	0.5	10 s	No	15.82 (11.76)	17.18 (15.14)	19.05 (11.45)	-	-	-	No	n/a
10	recognition	0.5	10 s	No	70.24 (16.54)	66.33 (17.37)	74.24 (15.32)	-	-	-	Yes 5>18	n/a
11A (JOLs condition)	recognition	0.5	10 s	Yes	84.52 (14.14)	82.60 (15.76)	80.94 (15.63)	49.71 (19.44)	45.35 (16.94)	36.21 (18.11)	No	Yes 48,18>5
11B (no-JOLs condition)	recognition	0.5	10 s	No	76.89 (15.61)	71.43 (17.50)	76.59 (14.89)	-	-	-	Yes 48>18 5>18	n/a

* non-contingent on reading; **Based on one-way Analysis of Variance and Bonferroni-adjusted post-hoc tests

point font, and words from the third set were presented in 5-point font. The assignment of font sizes to sets was counterbalanced across participants. In addition, three primacy buffers and three recency buffers were presented at the beginning and the end of the list respectively. The first, second, and third primacy and recency buffers always appeared in 48-, 18-, and 5-point font, respectively. Immediately after the presentation of each word, participants were prompted to predict the chance that they would later be able to remember it on a 0–100 scale (JOL). Participants were given 4 s to record their JOL on a form which included 48 empty fields labeled 1–48. They were instructed to write “x” in the appropriate field if they were unable to read the word. Immediately following the study list, participants engaged in a filler task for 5 min that required them to write down as many countries as they could. Finally, participants were given 4 min to write down as many of the study words as they could on a blank sheet of paper.

Results

The responses on the JOLs forms suggested that all participants were able to read all the words. Overall, participants correctly recalled 21% of the words, and JOLs averaged 49%. Table 1 presents descriptive statistics by font size for this and all subsequent experiments.

JOLs JOLs were significantly affected by font size, $F(2, 62) = 15.73$, $MSE = 134.93$, $p < .001$, $\eta_p^2 = .34$. Bonferroni-adjusted post-hoc tests revealed that JOLs were significantly ($p < .05$) higher for the 48-point words than for the 18-point words, and for the 18-point words than for the 5-point words.

Memory performance The percentage of words recalled was not significantly affected by font size, $F(2, 62) = .71$, $MSE = 96.54$, $p = .496$, $\eta_p^2 = .02$, and was similar for the 48-point, 18-point, and 5-point font words.

Experiment 2: Recognition, long study time, JOLs

Method

Participants Thirty-four students from the University of Haifa participated in the experiment. They were tested individually and received monetary compensation for their participation. Three participants were excluded from the analyses because of technical problems and one participant was excluded because he failed to return the test within 5 h of the study session. The final sample included 30 participants (24 women, age range: 19–34 years, mean age = 25.13 years).

Materials The same materials as in Experiment 1 were used for the study phase. An additional set of 42 words served as distractors in the recognition test. These words did not significantly differ from the study words with respect to familiarity, concreteness, number of letters, and number of syllables.

Procedure The procedure was similar to that of Experiment 1 except that: (a) a recognition test was used instead of a recall test, and (b) the retention interval was longer, because pretesting indicated that recognition performance would be at ceiling after a 5-min retention interval.

Procedure for the study phase was identical to that used in Experiment 1. At the end of the study phase participants were dismissed, and informed that the test would be sent to them by email in the next few hours. The test sheet was sent electronically to participants about 2 h after the study phase, and they were asked to complete it and send it back via email no later than 5 h after the study session. The test sheet included a list of all the studied words (excluding the primacy and recency buffers) and distractors (total of 84 words) presented in a fixed, random order. For each word, participants were asked to indicate whether it had appeared in the study phase or not.

Results

One participant was unable to read one 18-point word. Other than that, all participants were able to read all words. On average, participants returned the test about 3.5 h (213 min) after the study session (range 125–300 min). Overall, participants correctly recognized 85% of the studied words (hit rate), and falsely recognized 17% of the new words (false alarm rate). JOLs averaged 41%.

JOLs JOLs were significantly affected by font size, $F(2, 58) = 12.86$, $MSE = 118.87$, $p < .001$, $\eta_p^2 = .31$. Bonferroni-adjusted post-hoc tests revealed that JOLs for the 48-point words and 18-point words were significantly ($p < .05$) higher than for the 5-point words, but JOLs did not significantly differ between the 48-point and 18-point words.

Memory performance Non-contingent on successful reading, hit rates were not significantly affected by font size, $F(2, 58) = 1.38$, $MSE = 100.20$, $p = .260$, $\eta_p^2 = .05$, and were similar for the 48-point, 18-point, and 5-point words. Essentially the same results were obtained when hit rates were contingent on reading ($F(2, 58) = 1.39$, $MSE = 99.36$, $p = .26$, $\eta_p^2 = .05$; 48-point: $M = 87.62$, $SD = 9.37$; 18-point: $M = 85.38$, $SD = 12.40$; 5-point: $M = 83.33$, $SD = 11.46$).

Experiment 3: Free recall, short study time, JOLs

Method

Participants Thirty-two students from Yezreel Valley College participated in the experiment. They received either course credit or monetary compensation for their participation. Two participants were excluded because of technical problems, resulting in a final sample of 30 participants (19 women, age range: 21–29 years, mean age = 25.10 years).

Materials and procedure The materials and procedure were the same as in Experiment 1 except that (a) the words were presented for 0.5 s each and (b) the retention interval was 10 s, during which participants were asked to count backwards from 739, because pretesting with this short study time suggested that recall performance is at floor after 5 min (cf. Yue et al., 2013).

Results

Across all 30 participants and 42 items (1,260 cases) there were 30 cases (2.38%) in which participants failed to read the word (18, 9, and three cases for 5-, 18-, and 48-point words, respectively). Overall, participants correctly recalled 24% of the words and JOLs averaged 46%.

JOLs JOLs were significantly affected by font size, $F(2, 58) = 12.36$, $MSE = 92.71$, $p < .001$, $\eta_p^2 = .30$. Bonferroni-adjusted post-hoc tests revealed that JOLs were significantly ($p < .05$) higher for the 48-point words than for the 18-point words, and for the 18-point words than for the 5-point words.

Memory performance Non-contingent on reading, the percentage of words recalled was not significantly affected by font size, $F(2, 58) = .77$, $MSE = 129.12$, $p = .468$, $\eta_p^2 = .03$, and was similar for the 48-point, 18-point, and 5-point words. Essentially the same results were obtained when recall was contingent on reading ($F(2, 58) = .78$, $MSE = 138.02$, $p = .465$, $\eta_p^2 = .03$; 48-point: $M = 23.41$, $SD = 13.32$; 18-point: $M = 22.69$, $SD = 14.20$; 5-point: $M = 26.26$, $SD = 13.14$).

Experiment 4: Recognition, short study time, JOLs

Method

Participants Forty-three students from Yezreel Valley College and Tel-Aviv University participated in the experiment. Participants were tested individually and received either course credit or monetary compensation for their participation. One participant was excluded from the analyses because of technical problems. The final sample included 42 participants (32 women, age range: 22–30 years, mean age = 24.95 years).

Materials and procedure The procedure was the same as in Experiment 3, except that the test was a recognition test as in Experiment 2.

Results

Across all 42 participants and 42 items (1,764 cases) there were 43 cases (2.44%) in which participants failed to read the word (31, six, and six cases for 5-, 18-, and 48-point words, respectively). Overall, participants correctly recognized 82% of the studied words (hit rate), and falsely recognized 11% of the new words (false alarm rate). JOLs averaged 48%.

JOLs JOLs were significantly affected by font size, $F(2, 82) = 26.90$, $MSE = 149.51$, $p < .001$, $\eta_p^2 = .40$. Bonferroni-adjusted post-hoc tests revealed that JOLs were significantly ($p < .05$) higher for the 48-point words than for the 18-point words, and for the 18-point words than for the 5-point words.

Memory performance Non-contingent on reading, hit rates were not significantly affected by font size, $F(2, 82) = 2.17$, $MSE = 119.81$, $p = .120$, $\eta_p^2 = .05$, and were similar for the 48-point, 18-point, and 5-point words. Essentially the same results were obtained when recognition was contingent on reading ($F(2, 82) = 1.09$, $MSE = 111.62$, $p = .34$, $\eta_p^2 = .03$; 48-point: $M = 83.60$, $SD = 15.50$; 18-point: ($M = 82.54$, $SD = 15.54$; 5-point: $M = 80.26$, $SD = 17.37$).

Discussion of Experiments 1–4

Experiments 1–4 examined the effect of font size when JOLs were solicited and yielded two main findings. First, JOLs generally increased with font size. Second, memory performance was not affected by font size. These findings replicate the results of earlier studies and extend them as they were obtained even when a very small (5-point) font was included, regardless of test format and study time.

Experiments 5–8

Experiments 5–8 were designed to replicate Experiments 1–4, respectively, except that JOLs were not solicited. Since no data were collected during the study phase in these experiments (i.e., there was no JOLs form), there was no evidence on whether or not participants were able to read the words. However, the data from Experiments 1–4 suggested that cases in which participants were unable to read a word were relatively rare (1%, 74 out of 5,628 observations in Experiments 1–4) and that findings were essentially the same in the contingent and non-contingent analyses.

Experiment 5: Free recall, long study time, no JOLs

Method

Participants, materials and procedure Forty-two students from the Bar-Ilan University (30 women, age range: 20–28 years, mean age = 23.50 years) participated in the experiment. They received monetary compensation for their participation.

Materials and procedure The materials and procedure were the same as in Experiment 1 except that (a) JOLs were not solicited, and a 1-s inter-word interval was introduced instead, and (b) words were presented on a 22-in. wide screen and no chin rest was used. Importantly, 5-point was still the smallest legible font size.

Results

Overall, participants correctly recalled 21% of the words. The percentage of words recalled was not significantly affected by font size, $F(2, 82) = 1.63$, $MSE = 106.94$, $p = .203$, $\eta_p^2 = .04$, and was similar for the 48-point, 18-point, and 5-point words.

Experiment 6: Recognition, long study time, no JOLs

Method

Participants Forty-six students from the Bar-Ilan University participated in the experiment. They received monetary compensation for their participation. Seven participants were excluded from all analyses because they did not return the recognition test on time ($n = 2$; results are essentially the same when these participants are not excluded), did not return it at all ($n = 2$), or did not receive the test due to a technical problem ($n = 3$). The final sample included 39 participants (25 women, age range: 19–36 years, mean age = 25.05 years).

Materials and procedure The materials and procedure were the same as in Experiment 2 except that (a) as in Experiment 5, JOLs were not solicited and a 1-s inter-word interval was introduced instead, (b) as in Experiment 5, words were presented on a 22-in. wide screen and no chin rest was used, and (c) the recognition test was completed on-line using Qualtrics. A link to the test was sent to participants about 2 h after the study session, and they were asked to complete it within 2 h. In the test, the test words were presented one at a time, in a different random order for each participant.

Results

On average, participants returned the test about 2.75 h (163 min) after the study session (range: 120–270 min). Overall, participants correctly recognized 77% of the studied words

(hit rate), and falsely recognized 26% of the new words (false alarm rate). Hit rates were not significantly affected by font size, $F(2, 76) = 1.31$, $MSE = 110.33$, $p = .276$, $\eta_p^2 = .03$, and were similar for the 48-point, 18-point, and 5-point words.

Experiment 7: Free recall, short study time, no JOLs

Method

Participants, materials and procedure Forty-two students from Yezreel Valley College (34 women, age range: 21–42 years, mean age = 25.17 years) participated in the experiment. They received either course credit or monetary compensation for their participation. The materials and procedure were the same as in Experiment 3, except that JOLs were not solicited.

Results

Overall, participants correctly recalled 18% of the words. The percentage of words recalled was significantly affected by font size, $F(2, 82) = 3.12$, $MSE = 115.63$, $p = .049$, $\eta_p^2 = .07$. Bonferroni-adjusted post-hoc tests revealed that recall was significantly higher ($p < .05$) for the 48-point words than for the 18-point words, but did not differ between the 18-point and 5-point words or between the 48-point and 5-point words.

Experiment 8: Recognition, short study time, no JOLs

Method

Participants, materials and procedure Forty-two students from Yezreel Valley College (38 women, age range: 20–35 years, mean age = 23.95 years) participated in the experiment. They received course credit or monetary compensation for their participation. The materials and procedure were the same as in Experiment 4, except that JOLs were not solicited.

Results

Overall, participants correctly recognized 69% of the studied words (hit rate), and falsely recognized 32% of the new words (false alarm rate). Hit rates were significantly affected by font size, $F(2, 82) = 4.05$, $MSE = 170.20$, $p = .021$, $\eta_p^2 = .09$. Bonferroni-adjusted post-hoc tests revealed that hit rates were significantly higher ($p < .05$) for the 5-point words than for the 18-point words, but did not significantly differ between the 48-point words and the 5-point or 18-point words.

Discussion of Experiments 5–8

Experiments 5–8 examined the effect of font size when JOLs were not solicited. As far as can be ascertained, this is the first attempt to do so. Results of Experiments 5 and 6 revealed that

when study time was relatively long, font size did not affect memory even when a very small (5-point) font was included and regardless of test format. These findings again replicate the results of earlier studies and extend them to a situation in which JOLs are not solicited. In contrast, results of Experiments 7 and 8, in which study time was relatively short, revealed novel findings. In these experiments, font size affected memory, though in a different manner in each of the experiments.

In Experiment 7, free recall was better for the 48-point font than for the 18-point font. This finding is consistent with the results of a recent meta-analysis (Luna et al., *in press*), but it was not previously observed in a single study. In contrast, results of Experiment 8 provide the first evidence that small font size can improve rather than impair learning and be a desirable difficulty. When JOLs were not solicited and study time was short, presenting words in a very small (5-point) font increased recognition performance relative to presentation in the 18-point font. This finding is strikingly consistent with the results of Rosner et al. (2015), who reported that intense blurring improved word recognition under the same conditions (no JOLs, short study time).

Experiments 9–10

Font size affected memory in only two of Experiments 1–8, Experiments 7 and 8. The effect sizes were medium and medium-high, respectively. Experiments 9 and 10 sought to replicate Experiments 7 and 8, respectively, in order to examine the reliability of the results.

Experiment 9: Free recall, short study time, no JOLs

Method

Participants, materials and procedure Forty-two students from Bar-Ilan University (25 women, age range 18–45 years, mean age = 24.88 years) participated in the experiment. They received monetary compensation for their participation. The materials and procedure were the same as in Experiment 7, except that words were presented on a 22-in. wide screen and no chin rest was used.

Results and discussion

Overall, participants correctly recalled 17% of the words. The percentage of words recalled was not significantly affected by font size, $F(2, 82) = .88$, $MSE = 125.06$, $p = .417$, $\eta_p^2 = .02$, and was similar for the 48-point, 18-point, and 5-point words. Therefore, the effect of font size on recall that was found in Experiment 8 was not replicated in Experiment 9.

Experiment 10: Recognition, short study time, no JOLs

Method

Participants, materials and procedure Forty-two students from Bar-Ilan University (33 women, age range: 18–28 years, mean age = 22.48 years) participated in the experiment. They received monetary compensation for their participation. The materials and procedure were the same as in Experiment 8, except that words were presented on a 22-in. wide screen and no chin rest was used.

Results and discussion

Overall, participants correctly recognized 70% of the studied words (hit rate), and falsely recognized 18% of the new words (false alarm rate). Hit rates were significantly affected by font size, $F(2, 82) = 5.66$, $MSE = 116.30$, $p = .005$, $\eta_p^2 = .12$. Bonferroni-adjusted post-hoc tests revealed that hit rates were significantly higher ($p < .05$) for the 5-point words than for the 18-point words, but did not differ significantly between the 48-point words and the 5-point or 18-point words. Thus, Experiment 10 successfully replicated the findings of Experiment 8, confirming that very small font size can be a desirable difficulty for recognition memory, if JOLs are not solicited during study and the study time is short.

Experiment 11: Recognition, short study time, JOLs versus no JOLs

Experiment 11 was designed to examine directly the prediction that soliciting JOLs moderates the effect of font size on memory. It compared the effect of font size on word recognition performance using three font sizes and a relatively short study time in two conditions – one with JOLs and one without JOLs.

Method

Participants, materials and procedure Based on the effect size obtained in Experiment 10, a-priori power analysis suggested that the sample size required to detect an effect of font size in the no-JOLs condition with a power of .80 and assuming an alpha of 0.05 is 38. An attempt was made to double this sample size, per condition, to detect an interaction between effect size and JOL condition. This enabled collection of data from 144 participants before the end of the semester. The participants were students from Bar-Ilan University (112 women, age range: 19–46 years, mean age = 24.09 years). They received monetary compensation for their participation. Participants were randomly assigned to the *JOLs* and the *no-JOLs* conditions, which were essentially replications of

Experiments 4 and 8 (or 10), respectively. Words were presented on a 22-in. wide screen and no chin rest was used.

Results and discussion

The JOLs form of one participant was uninterpretable and this participant was excluded from the JOLs analyses. Across all 71 participants who provided JOLs data and 42 items (2,982 cases) there were 82 cases (2.75%) in which participants failed to read the word (54, 16, and 12 cases for 5-, 18-, and 48-point words, respectively). Overall, participants correctly recognized 79% of the studied words (hit rate), and falsely recognized 15% of the new words (false alarm rate). JOLs averaged 44%.

JOLs In the JOLs condition, JOLs were significantly affected by font size, $F(2, 140) = 21.21$, $MSE = 159.01$, $p < .001$, $\eta_p^2 = .28$. Bonferroni-adjusted post-hoc tests revealed that JOLs for the 48-point words and for the 18-point words were significantly ($p < .05$) higher than for the 5-point words, but that JOLs for the 48-point words and for the 18-point words did not differ significantly.

Memory performance Hit rates were analyzed in a 2 (condition: JOLs, no-JOLs) \times 3 (font size: 48-point, 18-point, 5-point) mixed-design analysis of variance. Data on whether participants failed to read a word were available only for the JOLs condition. Therefore, to avoid an item selection bias, the factorial analysis was conducted on hit rates that were non-contingent on reading. This analysis revealed a significant main effect of condition, $F(1, 142) = 12.78$, $MSE = 503.19$, $p < .001$, $\eta_p^2 = .08$, with higher hit rates in the JOLs ($M = 82.69$, $SD = 12.87$) than in the no-JOLs condition ($M = 74.97$, $SD = 13.03$). This finding is consistent with recent evidence that JOLs might improve memory (e.g., Soderstrom et al., 2015).

The analysis of hit rates also revealed a significant main effect of font size, $F(2, 284) = 4.29$, $MSE = 114.34$, $p = .015$, $\eta_p^2 = .03$. Bonferroni-adjusted post-hoc tests revealed that hit rates for the 48-point words ($M = 80.70$, $SD = 15.32$) were significantly ($p < .05$) higher than for the 18-point words ($M = 77.01$, $SD = 17.52$), but that hit rates for the 5-point words ($M = 78.77$, $SD = 15.36$) did not significantly differ from the other two font sizes. More importantly, the analysis also revealed a significant interaction between condition and font size, $F(2, 284) = 3.66$, $MSE = 114.34$, $p = .027$, $\eta_p^2 = .025$, suggesting that the JOL condition moderates the effect of font size, as predicted.

To interpret this interaction, the effect of font size on hit rates was analyzed separately for the JOLs and no-JOLs conditions. In the JOLs condition, hit rates numerically increased with font size, but this effect was not significant, $F(2, 142) = 2.36$, $MSE = 97.62$, $p = .098$, $\eta_p^2 = .032$. The effect of font size was also non-significant when hits were contingent on reading

($F(2, 142) = .97$, $MSE = 81.15$, $p = .382$, $\eta_p^2 = .013$; 48-point: $M = 85.31$, $SD = 13.53$; 18-point: $M = 83.36$, $SD = 15.30$; 5-point: $M = 83.52$, $SD = 14.43$).

In contrast, in the no-JOLs condition there was a significant effect of font size on hit rates, $F(2, 142) = 5.18$, $MSE = 131.05$, $p = .007$, $\eta_p^2 = .07$, replicating the results of Experiments 8 and 10. Bonferroni-adjusted post-hoc tests revealed a non-linear pattern in which hit rates for the 18-point words were significantly ($p < .05$) lower than for the 48-point words and for the 5-point words, but that hit rates for the 48-point words and 5-point words did not significantly differ from each other.

To supplement the analysis of hit rates, false alarm rates were analyzed. This analysis revealed a significant effect of condition, $t(142) = 6.04$, $p < .001$, Hedge's $g = 1.00$, with higher false alarm rates in the no-JOLs ($M = 20.55$, $SD = 12.19$) than in the JOLs condition ($M = 9.88$, $SD = 8.70$). The evidence that soliciting JOLs produced higher hit rates and lower false alarm rates suggests that it produced better discriminability rather than a mere criterion shift. To examine this conjecture, signal detection theory measures of sensitivity and response bias were calculated for each participant. Indeed, sensitivity (d') was higher in the JOLs ($M = 1.80$, $SD = .67$) than in the no-JOLs condition ($M = 1.18$, $SD = .53$), $t(142) = 6.06$, $p < .001$, Hedge's $g = 1.02$ (analysis of corrected hit rates yielded essentially the same results). However, response bias (c) was numerically higher (i.e., more conservative) in the JOLs ($M = .18$, $SD = .29$) than in the no-JOLs condition ($M = .09$, $SD = .32$), but this difference was not statistically significant, $t(142) = 1.69$, $p = .092$, Hedge's $g = .29$.

Summary of Experiments 1–11

Eleven experiments systematically investigated the effect of font size (large, small, and very small) on JOLs and memory performance, and yielded the following findings. First, JOLs generally increased with font size. Second, in Experiment 7, free-recall performance was better for the large (48-point) than the small (18-point) font when study time was relatively short and JOLs were not solicited, but this effect failed to replicate in Experiment 9. Third, recognition performance was better for the very small (5-point) than for the small (18-point) font when study time was relatively short and JOLs were not solicited. This effect was obtained in Experiment 8 and was replicated in Experiment 10 and in the no-JOLs condition of Experiment 11. In addition, in Experiment 11 recognition performance was better for the large (48-point) than for the small (18-point) font. Fourth, Experiment 11 suggested that solicitation of JOLs moderates the effect of font size on recognition memory. This effect is obtained only when JOLs are not solicited.

Meta-analyses

A series of small-scale meta-analyses was further conducted on the data from the reported experiments in order to examine whether font size was associated with memory beyond experiments and to assess the moderating role of solicitation of JOLs, study time, and test format that were mainly manipulated between experiments¹.

Since the JOLs and no-JOLs conditions in Experiment 11 were independent, they were treated as two separated studies labeled 11A and 11B, respectively, to allow a better assessment of the moderating role of solicitation of JOLs. The meta-analyses were therefore conducted on a set of 12 studies, with a total sample size of 527 participants.

Analyses were conducted separately for each font-size pairwise comparison (48-point vs. 18-point; 18-point vs. 5-point; and 48-point vs. 5-point). First a meta-analysis was conducted across all studies to test whether the effect size of font size was different from zero. Then, moderator (subgroup) analyses were conducted to assess the moderating role of solicitation of JOLs, study time, and test format. Note that these subgroup analyses were based on a relatively small number of studies in each subgroup and therefore results should be interpreted with caution. Because data on whether participants were able to read the words was available only for studies in which JOLs were solicited, the meta-analyses were conducted on memory performance that was non-contingent on reading, to avoid item-selection biases. However, results were essentially the same when the analyses were based on contingent memory performance for the studies that included JOLs.

Analyses were conducted using the Comprehensive Meta-Analysis software version 3.3.070 (Borenstein, Hedges, Higgins, & Rothstein, 2013) using the random effects model and Hedge's g as the measure of effect size. Positive effect sizes indicated a benefit of the larger over the smaller font size, and negative effect sizes – vice versa. Results are presented in Tables 2 and 3.

Forty-eight-point versus 18-point The meta-analysis revealed that memory performance was better for the 48-point words than for the 18-point words, Hedges's $g = .15$ ($SE = .05$), $Z = 2.99$, $p = .003$, replicating the finding of Luna et al. (2018). Heterogeneity between studies was low, $Q(11) = 12.87$, $p = .302$. Moderator analyses revealed that none of the potential moderators had an impact on the main effect: test format: $Q(1) = .22$, $p = .637$; study time: $Q(1) = .89$, $p = .344$; solicitation of JOLs: $Q(1) = .13$, $p = .717$.

Eighteen-point versus 5-point The meta-analysis revealed that memory performance was better for the 5-point words than for the 18-point words, Hedges's $g = -.17$ ($SE = .07$), $Z = -2.41$, $p = .016$. This finding is consistent with the idea that very small font size can be a desirable difficulty. Heterogeneity was significant, $Q(11) = 24.51$, $p = .011$. A moderator analysis revealed that solicitation of JOLs significantly moderated this effect, $Q(1) = 11.85$, $p = .001$. When analyzed separately, the benefit of the 5-point font over the 18-point font was obtained when JOLs were not solicited, Hedges's $g = -.29$ ($SE = .06$), $Z = -4.58$, $p < .001$, but not when JOLs were solicited, Hedges's $g = .03$ ($SE = .07$), $Z = .47$, $p = .640$. Moderator analyses further revealed that neither test format nor study time impacted the effect of font size, $Q(1) = .68$, $p = .410$ and $Q(1) = .09$, $p = .760$, respectively.

Forty-eight-point versus 5-point The meta-analysis revealed that font size was not associated with memory when comparing the 48-point and the 5-point words, Hedges's $g = -.02$ ($SE = .07$), $Z = -.21$, $p = .834$. However, heterogeneity was significant, $Q(11) = 29.27$, $p = .002$. A moderator analysis revealed that solicitation of JOLs moderated the effect, $Q(1) = 8.37$, $p = .004$. When JOLs were solicited, there was a benefit of the 48-point font over the 5-point font, Hedges's $g = .19$ ($SE = .09$), $Z = 2.17$, $p = .030$. When JOLs were not solicited, memory was better for the 5-point words than for the 48-point words, but this effect was only marginally significant, Hedges's $g = -.14$ ($SE = .07$), $Z = -1.92$, $p = .055$. This pattern merely reflects the effects that were obtained for the other two pairwise comparisons. When JOLs were solicited, there was a memory advantage for the 48-point over the 18-point font, and no difference between the 18-point and the 5-point fonts, hence there was a memory advantage for the 48-point over the 5-point font. When JOLs were not solicited, there was a memory advantage for the 48-point over the 18-point font, but also a memory advantage for the 5-point over the 18-point font, which yielded almost comparable levels of memory for the 48-point and the 5-point fonts. Moderator analyses further revealed that neither test format nor study time impacted the effect of font size, $Q(1) = .38$, $p = .535$ and $Q(1) = .04$, $p = .844$, respectively.

General discussion

Eleven experiments investigated the effect of font size (large, small, and very small) systematically under different conditions (with or without solicitation of JOLs, with free recall or recognition tests, and with relatively long or short study times and correspondingly longer or shorter retention intervals) and the data were meta-analyzed. Results are valuable for the understanding of the effect of font size on memory of words and metamemory, for demonstrating the moderating role of JOLs,

¹ For the sake of completeness, meta-analyses were also conducted on the JOLs data. To preview, these analyses revealed that JOLs consistently increased with font size. Full details appear in the Supplementary Materials.

Table 2 Effect sizes of font size for memory performance

Experiment	n	Test format	Study time / word (sec)	JOLs solicited?	48-pt versus 18-pt		18-pt versus 5-pt		48-pt versus 5-pt				
					Hedges's g	SE	95% CI	Hedges's g	SE	95% CI	Hedges's g	SE	95% CI
1	32	free recall	5	Yes	.15	.22	[-.29, .59]	-.23	.18	[-.58, .11]	-.09	.21	[-.51, .33]
2	30	recognition	5	Yes	.21	.26	[-.31, .73]	.15	.22	[-.29, .59]	.39*	.18	[.05, .74]
3	30	free recall	0.5	Yes	.08	.23	[-.37, .54]	-.26	.19	[-.64, .11]	-.18	.22	[-.61, .25]
4	42	recognition	0.5	Yes	.08	.15	[-.22, .39]	.20	.16	[-.11, .51]	.28*	.12	[.06, .51]
5	42	free recall	5	No	.23	.16	[-.08, .54]	-.28	.19	[-.65, .08]	-.04	.16	[-.34, .27]
6	39	recognition	5	No	-.11	.13	[-.37, .14]	-.12	.15	[-.41, .17]	-.23	.15	[-.52, .06]
7	42	free recall	0.5	No	.50*	.20	[.12, .89]	-.32	.22	[-.75, .12]	.17	.20	[-.22, .56]
8	42	recognition	0.5	No	.04	.16	[-.27, .35]	-.41*	.17	[-.73, -.08]	-.43*	.20	[-.82, -.04]
9	42	free recall	0.5	No	-.10	.20	[-.48, .29]	-.14	.17	[-.47, .20]	-.27	.19	[-.65, .10]
10	42	recognition	0.5	No	.23	.14	[-.05, .50]	-.47**	.16	[-.79, -.16]	-.25	.13	[-.51, .01]
11A JOLs	72	recognition	0.5	Yes	.13	.12	[-.10, .35]	.10	.09	[-.07, .28]	.24*	.12	[.01, .46]
11B no-JOLs	72	recognition	0.5	No	.32**	.11	[.11, .54]	-.31*	.12	[-.55, -.07]	.02	.13	[-.23, .27]
Mean					.15**	.05	[.05, .25]	-.17*	.07	[-.30, -.03]	-.02	.07	[-.16, .13]

Note. n = number of participants; SE = standard error; CI = confidence interval; * = $p < .050$; ** = $p < .010$; *** = $p < .001$; Positive effect sizes indicate a benefit of the larger over the smaller font; negative effect sizes indicate a benefit of the smaller over the larger font

and, more generally, for reconciling some of the recent inconsistency regarding the effects of perceptual degradation manipulations on learning outcomes.

The effect of font size on memory

The results yielded a u-shape relationship between font size and memory. Memory was better for the large than for the small font words. This finding emerged in the meta-analysis that pooled the data from all the experiments together, although it was significantly obtained only in two cases (Experiment 7 and the JOL condition in Experiment 11). The effect seems to be a stable one given that it converges with the evidence from a recent meta-analysis that was conducted on a different set of studies (Luna et al., *in press*). In addition, memory was better for the very small than for the small font words. This finding was obtained significantly in three of the experiments (Experiments 8, 10, and the no-JOL condition in Experiment 11) that used the same procedure (short study time, no JOLs, and recognition test), and also emerged in the meta-analysis that pooled the data from all the experiments together. Such a mnemonic benefit of small font size has not been demonstrated before as previous studies commonly did not use very small fonts. It suggests that very small fonts can be added to the growing list of perceptual degradation manipulations that constitute desirable difficulties.

At first glance, one might assume that this finding would have important practical implications. Today, reading is more and more often done from the screens of small electronic devices (e.g., smartphones) and the current results suggest that this might benefit cognitive performance. However, results of Experiment 11 and of the meta-analysis also suggested that the mnemonic benefit of very small font is not robust, as it was eliminated when learners provided JOLs. Extreme caution should therefore be exercised in applying these findings to real life situations. In particular, if the benefit of very small fonts is eliminated when learners provide JOLs, it might also be eliminated when they spontaneously monitor their learning or when they are involved in other tasks that promote elaborated processing. The effect of font size, or any other perceptual manipulation, on cognitive performance must be examined in the specific context that is of interest before it is used as a mean of enhancing learning in that context. Moreover, across experiments, the benefits of both the large and the very small fonts (compared to the intermediate ones) were relatively small, and therefore of questionable practical relevance.

At this stage, we can only speculate as to why there is a u-shape relationship between font size and memory. One possibility is that the same mechanism drives both the benefit of the large font and the benefit of the very small font. For example, both very large and very small fonts might be more salient and therefore remembered better than intermediate fonts (cf. Madan & Spetch, 2012; for related evidence that font size

Table 3 Moderator meta-analyses for memory performance

Moderator	<i>k</i>	48-pt versus 18-pt				18-pt versus 5-pt				48-pt versus 5-pt			
		Hedges's <i>g</i>	<i>SE</i>	95% CI	<i>Q_B</i>	Hedges's <i>g</i>	<i>SE</i>	95% CI	<i>Q_B</i>	Hedges's <i>g</i>	<i>SE</i>	95% CI	<i>Q_B</i>
JOLs solicitation					.13				11.85***				8.37**
JOLS	5	.12	.09	[-.05, .29]		.03	.07	[-.10, .17]		.19*	.09	[.02, .35]	
No JOLs	7	.16*	.06	[.04, .29]		-.29***	.06	[-.41, -.17]		-.14	.07	[-.28, .00]	
Test format					.22				.68				.38
Free recall	5	.19	.10	[-.00, .37]		-.24*	.11	[-.47, -.02]		-.08	.13	[-.33, .17]	
Recognition	7	.13*	.06	[.01, .25]		-.12	.09	[-.29, .05]		.02	.10	[-.17, .21]	
Study time					.89				.09				.04
Long	4	.07	.09	[-.11, .26]		-.13	.13	[-.39, .12]		.01	.14	[-.27, .28]	
Short	8	.18**	.06	[.06, .29]		-.18*	.09	[-.35, -.01]		-.03	.08	[-.17, .14]	

Note. *k* = number of studies; *SE* = standard error; CI = confidence interval; *Q_B* = heterogeneity between subgroups; * = *p* < .050; ** = *p* < .010; *** = *p* < .001; Positive effect sizes indicate a benefit of the larger over the smaller font; negative effect sizes indicate a benefit of the smaller over the larger font

effects on JOLs are driven by saliency, see Magreehan, Serra, Schwartz, & Narciss, 2016; Susser et al., 2013). Another possibility is that different mechanisms drive the benefit of the large font and the benefit of the very small font. For example, large fonts might be remembered better because font size is used as a proxy for importance and therefore larger fonts are processed better (Rhodes & Castel, 2008; Luna et al., in press). However, very small fonts might be remembered better regardless of perceived importance because they are more difficult to process, and the difficulty of processing triggers allocation of more cognitive resources and greater cognitive engagement (Diemand-Yauman et al., 2011; Hirshman & Mulligan, 1991). Examining these possibilities remains an issue for future research.

The moderating role of JOLs

As predicted, results suggest that the requirement to make JOLs moderates the mnemonic benefit of very small fonts. The evidence for the moderation emerged indirectly from the individual experiments that yielded this benefit only when JOLs were not solicited, consistent with similar prior demonstrations for perceptual interference and blurring (Besken & Mulligan, 2013; Rosner et al., 2015). Importantly, the current research also provides direct evidence for the moderating role of JOLs in the results of Experiment 11 and of the subgroup meta-analysis. These results suggest that previous studies that examined the effect of font size failed to observe a mnemonic benefit of smaller fonts, not only because they used font sizes that were not small enough to render this benefit, but also because they mainly focused on the effect on JOLs and followed Rhodes and Castel (2008) to include JOLs as part of their procedure. An important message of the current research is therefore that to identify the effect of a certain

manipulation on memory it is important to test it first without soliciting JOLs.

More research is needed to understand why JOLs moderate the benefit of very small fonts. One could argue that learners continue to process the words during the JOL interval, and that this additional processing masks the benefits of the small fonts that occur when JOLs are not solicited. However, the current results do not favor this explanation, because the meta-analysis suggested that study time does not moderate the benefit of very small fonts. Furthermore, Soderstrom et al. (2015) demonstrated that JOLs improve memory and moderate the effect of desirable difficulties (specifically, generation) even when the judgments are provided at the expense of some of the study time. A more plausible explanation might be that the processes that are involved in providing JOLs overlap with the processes that are responsible for the mnemonic benefits of perceptual manipulations (Rosner et al., 2015).

Two other potential moderators, test format and study time, were examined. Although in the individual experiments the benefit of the small font size was observed only for recognition and only when study time was short (and correspondingly, retention interval was short), the subgroup meta-analyses did not support the hypotheses that test format and study time moderate this benefit.

The effect of font size on JOLs

Although it was not the focus of the current research, it is also worth discussing how font size affected JOLs. In the current research, JOLs increased consistently with font size (see the Supplementary Materials for meta-analyses that examined whether font size was associated with JOLs), consistent with earlier studies (e.g., Rhodes & Castel, 2008). This finding was robust, as it persisted across different font sizes and paradigms. Interestingly, the results also yielded a complex

memory-JOLs relationship. When comparing the intermediate and large fonts, both memory and JOLs increased with font size, although the effect on JOLs was larger (consistent with Luna et al., [in press](#)). However, when comparing the intermediate and the very small fonts, a crossed double dissociation emerged: memory benefited from the smaller font but JOLs were lower. These results demonstrate the complex nature of the relationship between memory and JOLs. It is still debatable whether the higher JOLs for items presented in larger fonts are based on the subjective experience of relatively greater fluency (e.g., Susser et al., [2013](#); Yang, Huang, & Shanks, [2018](#)) or on a more general belief about the effect of font size on memory (Mueller et al., [2014](#)). Future research could examine the relative contributions of fluency and beliefs to the effect of very small font size on JOLs.

Conclusions

The results of the current research revealed that when learning words for a memory test, very small font size can be a desirable difficulty and hence the results provide support for the counterintuitive notion that perceptually degraded materials can enhance learning outcomes. More generally, results of the current research shed new light on the inconsistent findings of previous research on the effects of perceptual degradation on learning. They suggest that the null effect of perceptual manipulations in some of the previous studies might be attributed to manipulations that were not sufficiently strong to induce the processes that enhance memory. It would be interesting to examine whether procedures that yielded null effects of perceptual degradation on learning outcomes in previous studies (e.g., Eitel et al., [2014](#); Eitel & Kühl, [2016](#); Pieger et al., [2016](#); Rummer et al., [2016](#); Strukelj et al., [2016](#)) would yield an effect if more severely degraded materials were used. Moreover, the results suggest that inclusion of a JOL task in the study phase of many of the earlier studies may have masked the effect of perceptual degradation, and converge with previous findings (Besken & Mulligan, [2013](#); Rosner et al., [2015](#); Soderstrom et al., [2015](#)) to suggest that desirable difficulty effects may be difficult to observe when JOLs are solicited during learning. Collectively, the results of the current research suggest that caution should be exercised when generalizing the findings of research on memory and highlight the importance of systematic investigation of moderators.

Acknowledgements I am deeply grateful to Tami Katzir for her support during the initial stages of this research project. I also thank Liat Lev-Shalem for the statistical advice and the many research assistants who took part in data collection over the years. This research was partially supported by a seed money grant from the research authority of the University of Haifa.

References

- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576.
- Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory and Cognition*, *19*, 487–497.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68.
- Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, *41*(6), 897–903.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York: Worth Publishers.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2013). *Comprehensive meta-analysis* (Version 3.3.070). Englewood, NJ: Biostat.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671–684.
- Diemand-Yauman C, Oppenheimer D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, *118*, 114–118.
- Double, K. S., & Birney, D. P. (2017). Are you sure about that? Eliciting confidence ratings may influence performance on Raven's progressive matrices. *Thinking & Reasoning*, *23*(2), 190–206.
- Drori, S., & Henik, A. (2005). [Norms for examination of the concreteness effect]. In A. Henik, O. Rubinstein, & D. Anaki (Eds), *Norms for Hebrew words* (pp. 35–46). Beer-Sheeva, Israel: Ben-Gurion University of the Negev. (In Hebrew).
- Dunlosky, J., & Mueller, M. L. (2016). Recommendations for exploring the disfluency hypothesis for establishing whether perceptually degrading materials impacts performance. *Metacognition and Learning*, *11*(1), 123–131.
- Eitel, A., & Kühl, T. (2016). Effects of disfluency and test expectancy on learning with text. *Metacognition and Learning*, *11*, 107–121.
- Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: Does harder to read mean better to understand? *Applied Cognitive Psychology*, *28*(4), 488–501.
- French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, A., Heritage, J., Poland, R.L., Reece, S., & Tabor, J. (2013). Changing fonts in education: How the benefits vary with ability and dyslexia. *Journal of Educational Research*, *106*, 301–304.
- Halin, N., Marsh, J. E., Hellman, A., Hellström, I., & Sörqvist, P. (2014). A shield against distraction. *Journal of Applied Research in Memory and Cognition*, *3*(1), 31–36.
- Hirshman, E. & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 507–513.
- Hirshman, E., Trembath, D., & Mulligan, N. (1994). Theoretical implications of the mnemonic benefits of perceptual interference. *Journal*

- of *Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 608–620.
- Hu, X., Li, T., Zheng, J., Su, N., Liu, Z., & Luo, L. (2015). How much do metamemory beliefs contribute to the font-size effect in judgments of learning? *PLoS One*, 10(11), e0142351.
- Katzir, T., Hershko, S., & Halamish, V. (2013). The effect of font size on reading comprehension on second and fifth grade children: Bigger is not always better. *PLoS One*, 8, e74061.
- Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease of processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, 22, 787–794.
- Kühl, T., Eitel, A., Scheiter, K., & Gerjets, P. (2014). A call for an unbiased search for moderators in disfluency research: reply to Oppenheimer and Alter (2014). *Applied Cognitive Psychology*, 28(5), 805–806.
- Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning*, 11(1), 89–105.
- Lonsdale, M. D., Dyson, M. C., & Reynolds, L. (2006). Reading in examination type situation- the effects of text layout on performance. *Journal of Research in Reading*, 29, 433–453.
- Luna, K., Martín-Luengo, B., & Albuquerqu, P. B. (in press). Do Delayed Judgments of Learning Reduce Metamemory Illusions? A Meta-Analysis. *The Quarterly Journal of Experimental Psychology*.
- Madan, C. R., & Spetch, M. L. (2012). Is the enhancement of memory due to reward driven by value or salience?. *Acta Psychologica*, 139(2), 343–349.
- Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning*, 11(1), 35–56.
- McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–198). New York: Psychology Press.
- McDonough, I. M., & Gallo, D. A. (2012). Illusory expectations can affect retrieval-monitoring accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(2), 391–404.
- Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered? It depends on your beliefs about intelligence. *Psychological Science*, 22, 320–324.
- Miele, D. B., & Molden, D. C. (2010). Naive theories of intelligence and the role of processing fluency in perceived comprehension. *Journal of Experimental Psychology: General*, 139, 535.
- Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, 145(2), 200.
- Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory?. *Journal of Memory and Language*, 70, 1–12.
- Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1067–1087.
- Naime, J. S. (1988). The mnemonic value of perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 248–255.
- Naveh-Benjamin, M., & Kilb, A. (2012). How the measurement of memory processes can affect memory performance: The case of remember/know judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1), 194–203.
- Nguyen K., & McDaniel, M.A. (2016). The JOLs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied*, 22(1), 59–71.
- Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology*, 28(4), 502–504.
- Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency—Does disfluency lead to more accurate judgments, better control, and better performance?. *Learning and Instruction*, 44, 31–40.
- Rhodes, M. G. & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137, 615–625.
- Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica*, 160, 11–22.
- Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: null-effects of disfluency on learning outcomes. *Metacognition and Learning*, 11, 57–70.
- Schmidt, S. R., & Schmidt, C. R. (2017). Revisiting von Restorff's early isolation effect. *Memory & cognition*, 45(2), 194–207.
- Schnaubert, L., & Bodemer, D. (2017). Prompting and visualising monitoring outcomes: Guiding self-regulatory processes with confidence judgments. *Learning and Instruction*, 49, 251–262.
- Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science*, 45(2), 221–238.
- Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 553–558.
- Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: evidence from eye movements. *Metacognition and Learning*, 11, 71–88.
- Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin and Review*, 18, 973–978.
- Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory and Cognition*, 41(7), 1000–1011.
- Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgements of retention (JORs). *Quarterly Journal of Experimental Psychology*, 65(7), 1376–1396.
- Weissgerber, S. C., & Reinhard, M. A. (2017). Is disfluency desirable for learning?. *Learning and Instruction*, 49, 199–217.
- Weltman, D., & Eakin, M. (2014). Incorporating unusual fonts and planned mistakes in study materials to increase business student focus and retention. *INFORMS Transactions on Education*, 15(1), 156–165.
- Witherby, A. E., & Tauber, S. K. (2017). The influence of judgments of learning on long-term learning and short-term performance. *Journal of Applied Research in Memory and Cognition*, 6(4), 496–503.
- Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, 99, 99–110.
- Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is-and is not-a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, 41, 229–241.
- Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, 15, 41–44.