Psychological Research
https://doi.org/10.1007/s00426-022-01699-3

ORIGINAL ARTICLE

Revisiting the self-generation effect in proofreading

Alexander P. Burgoyne[1] · **Sari Saba-Sadiya**[2] · **Lauren Julius Harris**[2] · **Mark W. Becker**[2] · **Jan W. Brascamp**[2] · **David Z. Hambrick**[2]

Received: 23 January 2022 / Accepted: 6 June 2022

Abstract
The *self-generation effect* refers to the finding that people's memory for information tends to be better when they generate it themselves. Counterintuitively, when proofreading, this effect may make it more difficult to detect mistakes in one's own writing than in others' writing. We investigated the self-generation effect and sources of individual differences in proofreading performance in two eye-tracking experiments. Experiment 1 failed to reveal a self-generation effect. Experiment 2 used a studying manipulation to induce overfamiliarity for self-generated text, revealing a weak but non-significant self-generation effect. Overall, word errors (i.e., wrong words) were detected less often than non-word errors (i.e., misspellings), and function word errors were detected less often than content word errors. Fluid intelligence predicted proofreading performance, whereas reading comprehension, working memory capacity, processing speed, and indicators of miserly cognitive processing did not. Students who made more text fixations and spent more time proofreading detected more errors.

Introduction

Clarity and accuracy are essential in any kind of writing, especially in scientific reporting. Nevertheless, despite our best efforts, the occasional typographical error will go undetected, finding its way into a colleague's inbox or a submitted manuscript. These mistakes can be costly. Moreover, while proofreading is an important step in the writing process, sometimes proofreading our own work seems particularly difficult. The finding that it is more difficult to detect mistakes in one's own writing than in the writing of others, termed the self-generation effect in proofreading, is the focus of the current investigation.

Nearly forty years ago, Levy (1983) found that familiarity with a passage of text facilitated the detection of spelling errors. Participants read short passages and marked any errors that they noticed. In the *unfamiliar* condition, they simply proofread error-filled versions of the passages. In the *familiar* condition, they first read error-free versions multiple times and then proofread error-filled versions of the same passages. Participants in the familiar condition read faster and detected more errors than participants in the unfamiliar condition. Although the passages included only *non-word errors* (e.g., "about" changed to "ahout"), in subsequent work, Levy et al. (1986) found that familiarity also facilitated the detection of *word errors* (e.g., "major" changed to "mayor").

One possible explanation for these findings is that prior understanding of a passage allows readers to allocate more attention to error detection during proofreading. That is, familiarity makes reading comprehension more efficient, freeing mental resources for task-specific demands. As Levy et al. (1986) stated, "[P]rocessing efficiency is best viewed in terms of processes becoming faster and less resource demanding, so that more attention is available for strategic allocations within the task" (p. 488).

A straightforward implication of Levy and colleagues' results is that we should be excellent at proofreading our own writing because it is highly familiar. However, Daneman and Stainton (1993) found that students were *worse* at proofreading their own writing than others' writing. In their study, participants wrote a short essay on student life and were then assigned to one of three conditions. In the first condition, they proofread the essay they had just written. In the second condition, they proofread a different student's essay. In the third condition, they familiarized themselves with a different student's essay by reading it three

✉ Alexander P. Burgoyne
burgoyn4@gmail.com

1 School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

2 Department of Psychology, Michigan State University, East Lansing, MI, USA

times prior to proofreading an error-filled version. Everyone was told that errors had been added to the essays prior to proofreading.

Daneman and Stainton (1993) found that students who proofread their own essays detected 20% fewer errors than those who proofread a familiar essay written by someone else. By contrast, students who proofread a familiar essay written by someone else detected more errors than those who proofread an unfamiliar essay, replicating Levy et al.'s (1986) finding. This pattern of results suggests that familiarity enhanced proofreading but only up to a point—once an essay was *overfamiliar*, performance declined.

Supporting this interpretation, Daneman and Stainton (1993) next had participants proofread essays two weeks after writing them. They hypothesized that if overfamiliarity led to poor proofreading, then after two weeks participants' memory for their own essays would decay and overfamiliarity would no longer be a problem. Indeed, two weeks later, participants were able to proofread their own essays about as accurately as a familiar essay written by someone else.

Drawing on the idea of the self-generation effect, Daneman and Stainton (1993) argued that *overfamiliarity*, from having just written an essay and being "intimately acquainted with [its] semantic and syntactic features," hindered proofreading performance (p. 306). Specifically, they suggested that overfamiliarity caused participants to engage in a top-down, "expectancy-driven" (p. 299) style of processing that reduced attention to the visual and semantic-syntactic aspects of the text.

Today, nearly three decades later, Daneman and Stainton's (1993) hypothesis about the mechanism underlying the self-generation effect in proofreading has not been rigorously tested. Because eye movements are suggestive of the cognitive processing activities of readers (Rayner, 1998), eye tracking offers a potentially powerful way to test the self-generation effect. Specifically, eye movements could shed light on whether top-down, or "expectancy-driven," processing drives the relationship between overfamiliarity and poor proofreading.

Before discussing eye movements, however, it should be noted that the self-generation effect may not replicate. First, Daneman and Stainton's (1993) results were based on very small samples (e.g., $n = 10$ per condition in Experiment 1). Second, Pilotti and Chodorow (2009) conducted a larger study with multiple proofreading conditions and found that self-generated familiarity *facilitated* the detection of errors. These results are at odds with Daneman and Stainton (1993) but consistent with the literature showing that familiarity facilitates proofreading performance (e.g., Levy, 1983; Levy et al., 1986). Pilotti and Chodorow's (2009) method, however, was slightly different. For example, they had participants correct errors, whereas Daneman and Stainton (1993) did not. Still, their use of a larger sample calls into question the robustness of the self-generation effect in proofreading, and warrants further investigation.

Eye movements during reading

As Kintsch (2005) stated, "What we see is in part determined by what we expect to see" (p. 127). He argued that top-down and bottom-up processes interact to produce text comprehension and, in particular, suggested that *bottom-up processing* during reading—that is, a data-driven process based on analysis of the perceptual details of the text—results in knowledge activation. In turn, activated knowledge serves as a top-down constraint or backdrop against which further information is processed. From this standpoint, *top-down processing* reflects the influence of prior knowledge and expectations on perception.

Thus, a reader engaging primarily in bottom-up processing might fixate most words in a passage (Carrell & Eisterhold, 1983, pp. 567–568). Readers who fixate most words make shorter saccades because fixating every word necessitates short saccades between them. More bottom-up processing might also involve longer fixations because a thorough visual analysis requires more processing time (Rayner, 1998). In sum, readers engaging in more bottom-up processing might make more fixations, make shorter saccades, and have longer fixations than those engaging in less bottom-up processing.

Conversely, readers relying on top-down or "expectancy-driven" processing might not fixate some words in a passage, particularly if they are familiar with the text (Hyönä & Niemi, 1990). Instead, they might skip over words (i.e., making few fixations with large saccades in between), filling in the blanks with expectations about what is written (Pilotti & Chodorow, 2009, p. 247). They might also have shorter fixation durations if prior knowledge about the passage facilitates word recognition (Inhoff et al., 1993; Pilotti & Chodorow, 2009). Thus, readers engaging in more top-down processing might make fewer fixations, longer saccades, and shorter fixations than readers engaging in less top-down processing.

People typically fixate more words when reading unfamiliar text (Just & Carpenter, 1980). Presumably, unfamiliar text requires greater bottom-up processing because readers lack knowledge—a top-down influence—about it. Evidence for the influence of top-down processing on eye movements is provided by studies that have participants read the same passage multiple times. Subsequent readings are more conducive to top-down processing because knowledge of the text can be brought to bear during reading. For example, Hyönä and Niemi (1990) had participants read an essay twice for comprehension. On the second reading, participants made fewer fixations and regressions per sentence and had shorter fixation durations. Similar results were observed when participants read paraphrased text (Raney et al., 2000), suggesting that knowledge effects on reading behaviors generalize beyond the presentation of identical passages. Although these studies indicate that eye movements differ depending on the amount of knowledge a person has about the passage, it must be noted that the hypothesized link between patterns of eye movements and greater top-down or bottom-up processing is based on inferential evidence, not direct observation. For example, it is possible that other influences besides relying on top-down processing could cause readers with more knowledge about a passage to skip more words when reading it. We return to this point in the Discussion.

In general, evidence suggests that readers are less likely to fixate *function words* (e.g., and, of, the), which express grammatical relations between other words, than *content words* (e.g., essay, test, notes), which have compositional meaning (Just & Carpenter, 1980). This is because function words are typically shorter, higher in frequency, and more predictable, and these factors affect fixation probability (Paterson et al., 2020; Rayner, 1998; Schmauder et al., 2000). If function words are fixated less often, errors embedded in them might be particularly difficult to detect. Other studies (Levy et al., 1986; Schotter et al., 2014) have shown that word errors (i.e., wrong words) are harder to detect than non-word errors (i.e., misspellings), perhaps because the detection of word errors requires semantic processing, whereas non-words are erroneous regardless of the sentence they are embedded in.

Another set of factors that affect eye movements during reading are task demands. For example, Kaakinen and Hyönä (2010) presented participants with sentences and asked them to either proofread them or read them for comprehension. Participants made shorter intraword saccades and had longer fixation durations when proofreading compared to when reading for comprehension. They were also more likely to refixate on words when proofreading. In general, attentional resources appear to be modulated by task demands, possibly biasing participants towards greater bottom-up processing of orthographic features of text when proofreading.

These results were corroborated and extended by Schotter et al. (2014), who also compared the eye movements of participants when proofreading versus reading for comprehension. Schotter et al. (2014) found that participants spent more time when proofreading, and also had longer first fixation durations. They also were more likely to fixate target words when proofreading than when reading for comprehension. Schotter et al.'s (2014) results suggest that readers are able to flexibly prioritize different sub-component processes of reading depending on whether they are tasked with identifying errors or reading for comprehension.

Another study that bears on the present work examined satisfaction of search during proofreading (Barach et al., 2021). In non-proofreading visual search tasks (e.g., searching for a "T" among "L"s), the detection of a target stimulus typically reduces the likelihood that participants will detect a second target stimulus shortly thereafter. Barach et al. (2021) found that satisfaction of search also occurs during proofreading, such that after the detection of a typographical error, participants were less likely to detect a subsequent typo. Furthermore, participants engaged in expedited search, reducing their fixation durations and the number of refixations on the subsequent typo. This change in visual search behavior following successful target identification could reflect a global shift in strategy or attentional allocation during task performance, essentially reflecting task disengagement resulting from the sense that the task has been successfully accomplished. Although Barach et al. (2021) used word lists for their study, this tendency could pose a problem for naturalistic proofreading when it is unclear how many errors are embedded in the text.

Individual differences

Just as people differ in virtually all complex tasks, they probably differ in proofreading performance. If so, one factor that might account for at least some of this variance is cognitive ability, as it does for most complex tasks (Jensen, 1998). Daneman and Stainton (1993) found that reading comprehension scores correlated with proofreading ability. Presumably, individuals who were better able to understand reading material were better able to detect errors in writing. Working memory capacity, which reflects the ability to temporarily maintain and manipulate information, might also play a role in proofreading, given its relationship to reading comprehension (Arrington et al., 2014; Daneman & Carpenter, 1980). Fluid intelligence, which reflects novel problem-solving ability and pattern recognition, has also been shown to predict proofreading performance (Furnham, 2010; Furnham et al., 2006).

Dispositional and personality traits are also worth considering. Gallagher and Hall (1992) found a negative correlation between extraversion and proofreading performance which they attributed to hasty proofreading: "It may be that extraverts are impulsive and make this kind of error because they are rushing through the task, without giving enough thought to the task" (p. 234). Although Gallagher and Hall (1992) measured extraversion, their rationale is suggestive of "miserly cognitive processing," a term used by dual-process theorists to refer to individuals who tend to rely on *Type 1* automatic processing, providing "a quick solution that is a first approximation to an optimal response" (Toplak et al., 2014, p. 148). Cognitive misers might be more likely to rely on fast, automatic processes when proofreading, letting expectations override careful processing of the text and resulting in the detection of fewer errors. Thus, in the present study, we tested whether individual differences in these attributes predicted proofreading performance.

The present study

We investigated the self-generation effect in proofreading in two experiments. We extended previous work by using eye tracking during proofreading, and by collecting measures of ability and non-ability factors that could explain individual differences in proofreading performance. Experiment 1 was a near replication of Daneman and Stainton (1993), whereas Experiment 2 incorporated a studying manipulation to induce overfamiliarity. Our questions were: (1) Is there a self-generation effect in proofreading? (2) Is there an association between poor proofreading performance and top-down processing, as reflected by eye movements? (3) Are there differences in detection rates for word and non-word errors and for function and content errors? And (4) do individual-difference measures predict proofreading performance?

Experiment 1

Participants

The participants were 64 undergraduate students (53 women, $M_{age} = 19.28$, $SD = 1.29$) recruited from introductory psychology courses at Michigan State University. All reported normal or corrected-to-normal vision and stated that English was their first language. Of the 64 who attended Session 1, 50 returned one week later for Session 2. One participant did not follow instructions and was excluded from proofreading analyses. Another was dismissed from Session 2 because the eye tracker calibration failed. This left a useable sample of 63 for Session 1 and 48 for Session 2 for the proofreading analyses. Our power to detect an effect of the same magnitude ($d \approx 0.99$) as Daneman and Stainton (1993) was 0.97 for Session 1 and 0.92 for Session 2 (G*Power 3; Faul et al., 2007).

Procedure

Participants were tested individually. During Session 1, they typed a short essay on student life. Next, they completed a reading comprehension test while errors were added to their essays. After that, they proofread either their own essay or another participant's essay. Finally, they completed two cognitive ability tests, Raven's Matrices and Letter/Number Comparison.

Session 2 occurred one week later. First, participants completed another reading comprehension test. Next, they proofread either the essay they had written one week prior or another participant's essay. Finally, they completed the remaining cognitive ability tests.

Materials

Essay writing. Participants typed an essay on student life using a computer with "spell check" disabled. They were given the following instructions: "*In this task we would like you to write a short essay about college life. Try to write as quickly as possible because you will only be given 20 min. We would like you to write about three topics related to student life: your classes and coursework, food, and things students do for fun. Don't be concerned about the literary quality of your work or your typing accuracy. Your essay may be used to determine what events are most typical in a student's life.*" Participants typed until a large textbox presented on the screen was filled with text; essays were approximately 500 words long.

Adding errors to essays. First, an experimenter corrected any obvious errors in the essays, including spelling errors, basic grammatical errors, and punctuation errors. Next, a computer program added 20 errors to each essay: six function word errors (e.g., *or* changed to *of*), six function non-word errors (e.g., *are* changed to *ane*), four content word errors (e.g., *life* changed to *like*), and four content non-word errors (e.g., *notes* changed to *nofes*). Thus, 10 word errors and 10 non-word errors were added: 12 function errors and 8 content errors. The errors were randomly drawn from a list of target words to ensure that similar types of errors were embedded. The target word list was developed during pilot testing and included words that frequently occurred in participants' essays and the 24 errors provided in an example essay by Daneman and Stainton (1993, p. 303). Errors were never added to the first 10 words of an essay. The program generated up to 20 versions of each essay with errors randomly embedded and selected the version that maximized the number of words between errors.

Reading comprehension (Brown et al., 1993). Participants completed computerized versions of the Nelson-Denny Reading Test, different versions during Sessions 1 and 2. Each version included seven reading passages and 38 questions with five answer choices. The time limit was 20 min (Cronbach's alpha [α] = .72).

Proofreading task. Participants performed the proofreading task using a specially designed computer program.[1] They read the following instructions: "*In this task you will read an essay. A number of errors have been inserted into the essay. Your task is to read the essay and highlight the errors. You*

[1] We thank David MacFarlane for his help developing the proofreading program.

can highlight errors in the essay by clicking and dragging with the left mouse button. To unhighlight something, click and drag with the right mouse button. You should try to proofread as quickly and accurately as possible." Participants were told to press the "escape" key on the keyboard when they finished proofreading the essay.

Participants were randomly assigned to one of two conditions. In Session 1, participants in the *self-generated* condition proofread the essay that they had written 20 min prior, whereas participants in the *other-generated* condition proofread an essay written by another participant. Participants were tested individually, but their condition assignments were yoked in pairs, such that the first participant in each pair was assigned to the self-generated condition and the second participant was assigned to the other-generated condition and proofread the essay written by the first participant in the pair. In Session 2, participants originally in the self-generated condition proofread the essay written by the other participant in the pair. Conversely, participants originally in the other-generated condition proofread the essay they had written one week prior.

Eye tracking measures. Eye movements were recorded using an EyeLink 1000 (SR Research Ltd., 2010) during the proofreading task. Left-eye gaze was tracked at 1000 Hz. Stimuli were presented on a monitor 692 mm from the participants' chin rest, with screen dimensions of 433 × 271 mm and a resolution of 1680 × 1050 px. Drift correction was administered prior to displaying the essay to be proofread. The EyeLink 1000 on-line parser was used to detect saccade and blink events. Additional filtering removed saccades larger than half the display size to omit eye movements between new lines.

We report three measures from the eye-tracking data. "Number of fixations" refers to the number of fixations made to the text during proofreading. "Saccade amplitude" refers to the mean distance of saccades made during proofreading, measured in pixels. "Fixation duration" refers to the mean duration of all fixations made while proofreading, measured in milliseconds.

Raven's advanced progressive matrices (Raven & Court, 1998). Participants were shown a set of patterns arranged in a 3 × 3 formation with the lower-right pattern missing. They selected from a list of options the pattern that logically completed the set. They were given 10 min to complete the 18 odd-numbered items ($\alpha = .76$).

Letter/number comparison (Salthouse & Babcock, 1991). Participants determined whether two strings of letters or numbers were the same or different. They were given 30 s for each set of 72 items; there were two sets of letter items and two sets of number items (split-half reliability = .78).

Symmetry span (Oswald et al., 2015). Participants were asked to judge whether black and white geometric designs were or were not symmetric while memorizing the position of squares appearing after each judgment. The measure was the number of correctly recalled square positions ($\alpha = .75$).

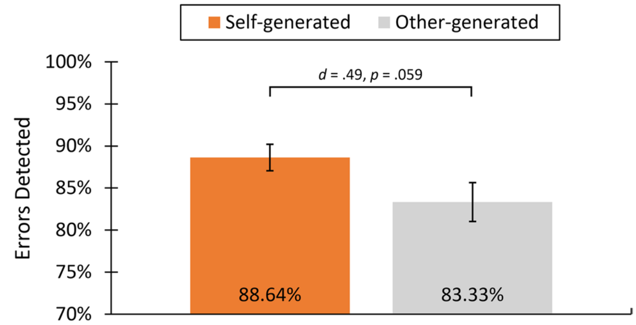

Fig. 1 Proofreading performance in Session 1 of Experiment 1. For all bar graphs, error bars represent ± 1 standard error around the mean

Pattern comparison (Salthouse & Babcock, 1991). Participants determined whether two symbols made from simple line drawings were the same or different. They were given 30 s for each set of 30 items; there were two sets of items. The measure was the number correct minus two times the number incorrect (split-half reliability = .68).

Visual arrays (Burgoyne et al., 2019). Participants were sequentially shown a memory array of two to eight colored squares, a blank display, and a test array, which was either identical to the memory array or different. Participants determined whether the arrays were the same or different. There were 80 trials ($\alpha = .93$).

Letter sets (Ekstrom et al., 1976). Participants were shown five sets of four capital letters (e.g., DEFG) and chose the set that did not follow the same pattern as the other four. Participants were given five minutes to complete 20 items ($\alpha = .76$).

Cognitive ability composites. Composite variables were formed by averaging *z*-scores (standardized scores). There were four composite variables: Reading Comprehension (Nelson-Denny Forms G and H), Working Memory Capacity (Symmetry Span, Visual Arrays), Fluid Intelligence (Raven's Matrices, Letter Sets), and Perceptual Speed (Letter/Number Comparison, Pattern Comparison).

Results of Experiment 1

Session 1 results

Based on Daneman and Stainton's (1993) results, we predicted that the self-generated group would detect fewer errors than the other-generated group during Session 1, but the results were in the opposite direction: the self-generated group detected 5.31% more errors than the other-generated group, although this difference was not

Table 1 Proofreading performance in Session 1 of Experiment 1

Measure	Self-generated	Other-generated	Difference between groups
Session 1 proofreading			
Total errors detected (%)	88.6 (09.0)	83.3 (12.7)	$t(61)=1.92$, $d=0.49$, $p=.059$
Word errors (%)	89.4 (09.3)	80.0 (20.0)	$t(61)=2.43$, $d=0.61$, $p=.018$
Non-word errors (%)	87.9 (14.3)	86.7 (11.2)	$t(61)=0.37$, $d=0.09$, $p=.712$
Content errors (%)	92.1 (09.8)	85.8 (15.0)	$t(61)=1.97$, $d=0.50$, $p=.054$
Function errors (%)	86.4 (11.0)	81.7 (15.8)	$t(61)=1.38$, $d=0.35$, $p=.173$
Session 1 eye tracking			
Fixations	660.9 (128.0)	776.7 (257.0)	$t(61)=2.30$, $d=0.58$, $p=.025$
Saccade amplitude (px)	44.5 (10.5)	38.4 (11.4)	$t(61)=2.21$, $d=0.56$, $p=.031$
Fixation duration (ms)	284.7 (30.8)	275.2 (37.3)	$t(61)=1.10$, $d=0.28$, $p=.274$
Time spent (s)	187.8 (38.4)	209.4 (58.7)	$t(61)=1.75$, $d=0.44$, $p=.085$

Means are presented with *SD*s in parentheses. Self-Generated $N=33$, Other-Generated $N=30$

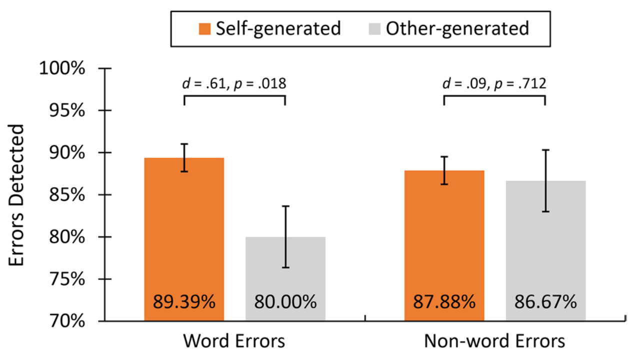

Fig. 2 Word and non-word error detection rates in Session 1 of Experiment 1

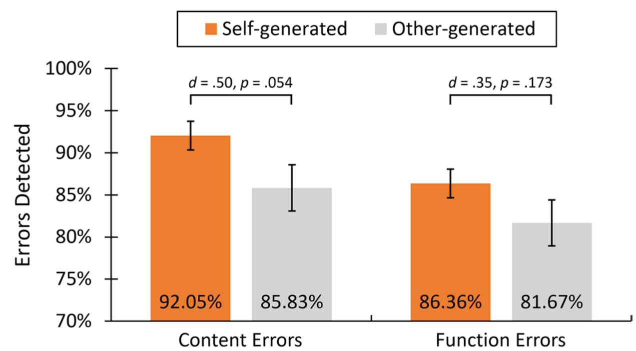

Fig. 3 Content and function error detection rates in Session 1 of Experiment 1

statistically significant, $t(61)=1.92$, $d=0.49$, $p=.059$ (Fig. 1, Table 1). Thus, Daneman and Stainton's (1993) results failed to replicate.

Word and non-word errors (Fig. 2). Participants detected slightly more non-word errors ($M=87.30\%$, $SD=12.85\%$) than word errors ($M=84.92\%$, $SD=15.95\%$), but this difference was not significant, $t(62)=1.02$, $d=0.16$, $p=.310$. The self-generated group detected significantly more word errors than the other-generated group, $t(61)=2.43$, $d=0.61$, $p=.018$. That is, familiarity from having generated the text oneself facilitated the detection of word errors. By comparison, the self-generated group did not detect significantly more non-word errors than the other-generated group, $t(61)=0.37$, $d=0.09$, $p=.712$.

Content and function errors (Fig. 3). Participants detected significantly more content errors ($M=89.09\%$, $SD=12.80\%$) than function errors ($M=84.13\%$, $SD=13.61\%$), $t(62)=2.67$, $d=0.38$, $p=.010$. The self-generated group detected 6.22% more content errors than the other-generated group, but this difference was not significant, $t(61)=1.97$, $d=0.50$, $p=.054$. The self-generated group did not detect significantly more function errors than the other-generated group, $t(61)=1.38$, $d=0.35$, $p=.173$.

Eye tracking. The self-generated group made significantly fewer fixations ($t(61)=2.30$, $d=0.58$, $p=.025$) and larger saccades ($t(61)=2.21$, $d=0.56$, $p=.031$) than the other-generated group (Table 1). Although the difference was not significant, the self-generated group spent less time proofreading than the other-generated group ($t(61)=1.75$, $d=0.44$, $p=.085$). The self-generated group and other-generated group did not differ in mean fixation duration, ($t(61)=1.10$, $d=0.28$, $p=.274$).

Number of fixations ($r=.27$, $p=.035$) and time spent proofreading ($r=.35$, $p=.005$) correlated significantly with proofreading performance (Table 2). Participants who made more fixations or spent more time proofreading detected more errors than those who made fewer fixations or spent less time proofreading.

Cognitive ability. Only fluid intelligence correlated significantly with overall proofreading performance ($r=.26$, $p=.043$; Table 2). In contrast to Daneman and Stainton (1993), the correlation between reading comprehension and proofreading performance was near zero ($r=.01$, $p=.950$).

Table 2 Correlations between proofreading performance, eye movements, and cognitive ability in Experiment 1

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.
1. Proofreading (S1)	–																				
2. Word errors (S1)	**.83**	–																			
3. Non-word errors (S1)	**.71**	.19	–																		
4. Content errors (S1)	**.74**	**.56**	**.58**	–																	
5. Function errors (S1)	**.91**	**.77**	**.61**	**.38**	–																
6. Proofreading (S2)	**.37**	**.45**	.08	.18	**.39**	–															
7. Word errors (S2)	**.33**	**.43**	.03	.15	**.35**	**.86**	–														
8. Non-word errors (S2)	.24	.24	.11	.13	.24	**.65**	.19	–													
9. Content errors (S2)	**.31**	**.34**	.10	.14	**.33**	**.63**	**.48**	**.51**	–												
10. Function errors (S2)	.28	**.36**	.04	.14	**.29**	**.88**	**.80**	**.52**	.18	–											
11. Fixations (S1)	**.27**	.14	**.29**	.16	**.27**	− .10	− .09	− .06	− .15	− .03	–										
12. Saccade amplitude (S1)	− .17	− .03	**− .26**	− .12	− .16	.09	− .01	.18	.19	− .01	**− .71**	–									
13. Fixation duration (S1)	.17	.19	.06	.16	.13	.14	.08	.15	.05	.14	**− .37**	.20	–								
14. Time spent (S1)	**.35**	.23	**.32**	.23	**.33**	− .03	− .04	.00	− .13	.05	**.89**	**− .68**	.08	–							
15. Fixations (S2)	.01	.06	− .05	− .02	.03	.27	.26	.13	.06	**.30**	**.49**	− .21	**− .35**	**.39**	–						
16. Saccade amplitude (S2)	− .04	− .07	.02	− .08	.00	− .24	− .24	− .11	− .04	− .28	**− .30**	**.30**	.15	− .27	**− .80**	–					
17. Fixation duration (S2)	.07	.02	.11	.22	− .03	.12	.12	.06	.07	.11	− .19	.00	**.75**	.13	− .21	.00	–				
18. Time spent (S2)	.02	.04	− .01	.06	.00	**.30**	**.29**	.14	.07	**.33**	**.40**	− .23	.01	**.45**	**.87**	**− .78**	.28	–			
19. Reading comprehension	.01	.08	− .09	.02	.00	.20	.10	.23	.18	.14	.15	− .08	− .21	.06	.06	.00	**− .30**	− .11	–		
20. WMC	.04	− .05	.13	.05	.02	.03	.07	− .04	− .11	.11	.16	− .21	− .01	.18	.21	− .18	.04	.22	.14	–	
21. Gf	**.26**	.21	.18	.04	**.33**	**.41**	.26	**.40**	**.36**	**.29**	.15	− .06	.15	.24	.09	.02	− .14	− .02	**.36**	**.30**	–

Table 2 (continued)

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.
22. PS	− .03	− .07	.03	− .01	− .03	− .18	− .12	− .17	− .20	− .10	.01	− .04	− .04	− .01	.04	− .02	− .04	− .01	.08	.20	− .02

Bolded correlations are significant at $p < .05$. *Ns* range from 48 to 63

S1 Session 1, *S2* Session 2, *WMC* working memory capacity, *Gf* fluid intelligence, *PS* processing speed

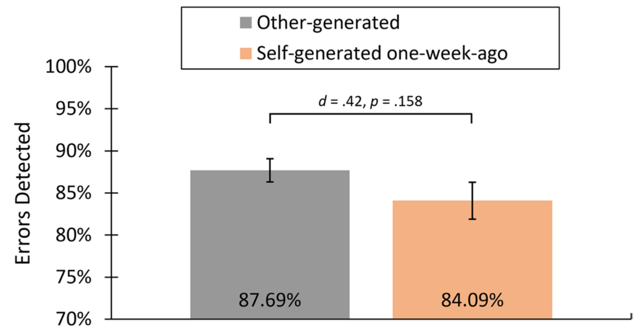

Fig. 4 Proofreading performance in Session 2 of Experiment 1

Working memory capacity ($r = .04$, $p = .791$) and processing speed ($r = -.03$, $p = .805$) were not significant predictors of proofreading performance.

Session 2 results

The names of the groups change from Session 1 to Session 2 to reflect the proofreading task during Session 2. The "self-generated" group from Session 1 is now called the "other-generated" group because in Session 2 they proofread someone else's essay. The "other-generated" group from Session 1 is called the "self-generated one-week-ago" group because in Session 2 they proofread the essay they wrote one week prior.

Session 2 revealed further lack of support for Daneman and Stainton's (1993) findings (Fig. 4, Table 3). Their results would predict that the self-generated one-week-ago group would outperform the other-generated group. Instead, we found no significant difference between them, $t(46) = 1.44$, $d = 0.42$, $p = .158$.

Word and non-word errors (Fig. 5). Participants detected significantly more non-word errors ($M = 90.83\%$, $SD = 8.95\%$) than word errors ($M = 81.25\%$, $SD = 13.47\%$), $t(47) = 4.51$, $d = 0.83$, $p < .001$. This difference is in the same direction as in Session 1, but only in Session 2 was it significant. The other-generated group did not detect significantly more word errors than the self-generated one-week-ago group, $t(46) = 0.16$, $d = 0.05$, $p = .874$. However, the other-generated group detected significantly more non-word errors, $t(46) = 2.70$, $d = 0.78$, $p = .010$.

Content and function errors (Fig. 6). Participants detected significantly more content errors ($M = 90.11\%$, $SD = 10.62\%$) than function errors ($M = 83.33\%$, $SD = 11.53\%$), $t(47) = 3.32$, $d = 0.61$, $p = .002$. This is consistent with Session 1. The other-generated group did not detect significantly more content errors than the self-generated one-week-ago group, $t(46) = 1.24$, $d = 0.36$, $p = .223$. The other-generated group did not detect significantly more function errors

Table 3 Proofreading performance in Session 2 of Experiment 1

Measure	Other-generated	Self-generated one-week-ago	Difference between groups
Session 2 proofreading			
Total errors detected (%)	87.7 (07.0)	84.1 (10.3)	$t(46)=1.44$, $d=0.42$, $p=.158$
Word errors (%)	81.5 (11.9)	80.9 (15.4)	$t(46)=0.16$, $d=0.05$, $p=.874$
Non-word errors (%)	93.8 (05.7)	87.3 (10.8)	$t(46)=2.70$, $d=0.78$, $p=.010$
Content errors (%)	91.8 (08.6)	88.1 (12.5)	$t(46)=1.24$, $d=0.36$, $p=.223$
Function errors (%)	84.9 (10.0)	81.4 (13.1)	$t(46)=1.04$, $d=0.30$, $p=.302$
Session 2 eye tracking			
Fixations	658.1 (183.6)	710.3 (212.4)	$t(46)=0.91$, $d=0.26$, $p=.366$
Saccade amplitude (px)	44.9 (12.7)	41.6 (11.2)	$t(46)=0.93$, $d=0.27$, $p=.356$
Fixation duration (ms)	265.9 (25.3)	277.2 (46.6)	$t(46)=1.06$, $d=0.31$, $p=.293$
Time spent (s)	174.4 (46.5)	194.5 (56.6)	$t(46)=1.35$, $d=0.39$, $p=.182$

Means are presented with *SD*s in parentheses. Self-Generated $N=26$, Other-Generated $N=22$

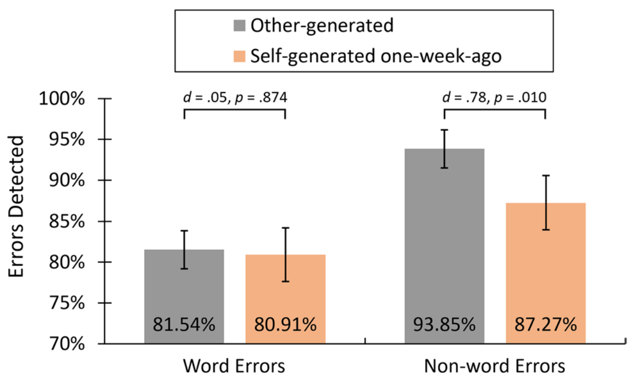

Fig. 5 Word and non-word error detection rates in Session 2 of Experiment 1

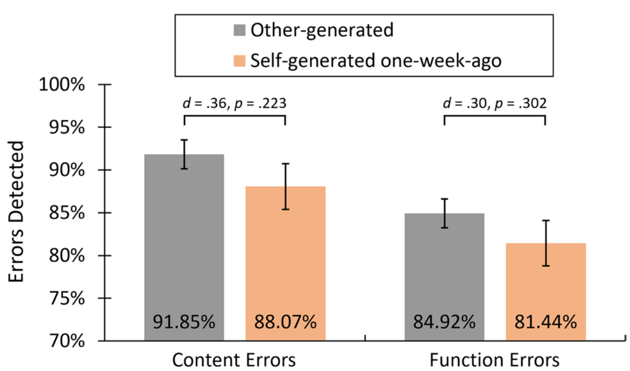

Fig. 6 Content and function error detection rates in Session 2 of Experiment 1

than the self-generated one-week-ago group, $t(46)=1.04$, $d=0.30$, $p=.302$.

Eye tracking. Although not statistically significant, the self-generated one-week-ago group made slightly more fixations, shorter saccades, had longer fixation durations, and spent more time proofreading than the other-generated group (all $ps>.18$; see Table 3).

As in Session 1, time spent proofreading correlated significantly with overall proofreading performance ($r=.30$, $p=.042$; Table 2). However, unlike in Session 1, the correlation between number of fixations and proofreading performance was non-significant ($r=.27$, $p=.064$).

Cognitive ability. As in Session 1, fluid intelligence correlated positively with proofreading performance ($r=.41$, $p=.004$; Table 2), whereas the correlation between reading comprehension and proofreading performance was non-significant ($r=.20$, $p=.183$). Working memory capacity ($r=.03$, $p=.833$) and processing speed ($r=-.18$, $p=.229$) were not significant predictors of proofreading performance.

Interim discussion

Experiment 1 did not replicate Daneman and Stainton's (1993) results. In Session 1, the self-generated group detected 5.3% more errors than the other-generated group ($p=.059$), a result which trends in the opposite direction as Daneman and Stainton's (1993) findings but is consistent with Pilotti and Chodorow's (2009) results. One post hoc explanation for this finding is that participants in the self-generated group were merely *familiar* with their essays, not *overfamiliar* with them. From this standpoint, the results are broadly aligned with previous research which has found that familiarity facilitates proofreading and is associated with less thorough visual processing of the text (Hyönä & Niemi, 1990; Levy et al., 1986). We reasoned that we may need to induce overfamiliarity for self-generated text to induce a self-generation effect (i.e., worse performance when proofreading one's own writing). To test this possibility, in Experiment 2 we incorporated a studying manipulation intended to induce greater familiarity with a passage of text.

Experiment 2

In Experiment 2, we attempted to produce the self-generation effect by inducing overfamiliarity of self-generated text. The manipulation was that some participants studied an essay prior to proofreading. Specifically, participants were assigned to a self-generated condition or an other-generated condition, as in Experiment 1, but they also were assigned to either study an essay prior to proofreading or not. Thus, there were four conditions: "self-study," "self-no study," "other-study," and "other-no study."

The major prediction was that studying would impair proofreading in the self-generated condition but facilitate proofreading in the other-generated condition. The rationale was that participants in the self-study condition would become "overfamiliar" with their essays and fail to detect errors in them. By contrast, participants in the self-no study condition would remain merely "familiar" with their essays, facilitating proofreading. A secondary prediction was that participants in the other-study condition would proofread better than participants in the other-no study condition, as this would be consistent with Levy and colleagues' findings that familiarity with a passage written by someone else facilitates proofreading performance.

For eye movements, the prediction was that as essay familiarity increased, so would hypothesized indicators of top-down processing. Therefore, we predicted that participants in the self-study condition (the "most familiar" group) would have the fewest fixations, largest saccades, shortest fixation durations, and spend the least time proofreading, and participants in the other-no study condition (the "least familiar" group) would have the most fixations, shortest saccades, the longest fixation durations, and spend the most time proofreading.

In Experiment 2, we also had participants report their ACT scores, college GPA, and complete the cognitive reflection test and need-for-cognition questionnaire. We tested whether these individual-difference measures predicted proofreading performance.

Methods

Participants

The participants were 100 undergraduate students (74 women, $M_{age} = 18.79$, $SD = 0.87$). We had an estimated power of .96 to detect a difference of $d = 0.99$ between two groups (Faul et al., 2007), the approximate magnitude of the difference observed by Daneman and Stainton (1993). Everyone reported normal or corrected-to-normal vision and stated that English was their first language.

Procedure

Participants first typed a short essay. Next, they completed a reading comprehension test. Those in the "study" conditions then spent five minutes reading an essay and committing it to memory. Afterward, everyone completed a memory test to assess essay familiarity. Next, they proofread either their own essay or another participant's essay. Finally, they completed a demographic questionnaire, the need-for-cognition questionnaire, and the cognitive reflection test.

Materials

Essay writing task. See Experiment 1.

Adding errors to essays. See Experiment 1.

Eye tracking measures. See Experiment 1.

Reading comprehension. This test was used in Session 1 of Experiment 1 ($\alpha = .83$).

Study time. Participants were randomly assigned to either study an essay or not, with the constraint that an equal number of participants were assigned to each condition. Participants in the self-study condition studied their own essay; participants in the other-study condition studied another participant's essay. The essay they studied was the essay they would later proofread. The essays had been corrected by the experimenter but were without added errors so as not to spoil the proofreading task. Participants were instructed: "*Please read the following essay three times and commit as much of it to memory as possible. There will be a memory test on this essay later on. You will have 5 min, starting now.*"

Memory test. Participants were shown 10 sentences and were asked whether they had seen each one before. Five had previously been shown; the remainder were foils taken from essays in Experiment 1. All participants were shown sentences from the essays they would later proofread, except for those in the other-no study condition: the sentences were those they had written, *not* sentences from the essay they would proofread. This is because the other-no study condition required participants to proofread an unfamiliar essay, and using sentences from the to-be-proofread essay might induce familiarity.

Proofreading task. See Experiment 1. Participants in the "self" conditions proofread their own essay and participants in the "other" conditions proofread another participant's essay.

Demographic questionnaire. Participants reported their age, sex, and GPA and ACT scores.

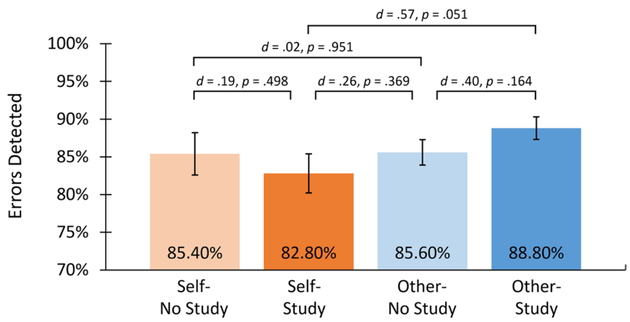

Fig. 7 Proofreading performance in Experiment 2

Need for cognition (Cacioppo et al., 1984). Participants completed the 18-item need-for-cognition questionnaire, designed to reflect the tendency for an individual to engage in and to enjoy thinking. An example item is "I only think as hard as I have to" (reverse scored). They rated their agreement with each item on a 5-point Likert scale ($\alpha = .89$).

Cognitive reflection test (Frederick, 2005). Participants completed the 3-item cognitive reflection test to measure their ability/disposition to consider a question and to inhibit reporting the first response that springs to mind. An example item is "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?" ($\alpha = .78$).

Results of Experiment 2

Consistent with prediction, the self-study group performed worst on the proofreading task, whereas the other-study group performed best (Fig. 7, Table 4). However, the group difference (6.0%) between the self-study and other-study conditions was not significant ($d = 0.57$, $p = .051$). We conducted a two-way ANOVA on proofreading performance, with essay condition (self vs. other) and study condition (study vs. no study) as between-subjects factors. The effect of essay condition was not significant ($F(1, 96) = 1.96$, $\eta_p^2 = 0.020$, $p = .165$); nor was the effect of study condition ($F(1, 96) = 0.02$, $\eta_p^2 = 0.000$, $p = .893$). The Essay Condition × Study Condition interaction was not significant ($F(1, 96) = 1.71$, $\eta_p^2 = 0.018$, $p = .194$).

Word vs. non-word errors. Participants detected significantly more non-word errors ($M = 88.20\%$, $SD = 11.40$) than word errors ($M = 83.20\%$, $SD = 15.03\%$), $t(99) = 3.30$, $d = 0.37$, $p = .001$. This result is consistent with Experiment 1.

Word errors (Fig. 8). The self-no study group detected the fewest word errors, whereas the other-study group detected the most. A two-way ANOVA with word error detection as the dependent measure revealed that the effect of essay condition was not significant ($F(1, 96) = 2.60$, $\eta_p^2 = 0.026$, $p = .110$), nor was the effect of study condition ($F(1, 96) = 1.46$, $\eta_p^2 = 0.015$, $p = .230$), or the Essay Condition × Study Condition interaction ($F(1, 96) = 0.88$, $\eta_p^2 = 0.009$, $p = .349$).

Non-word errors (Fig. 9). The self-study group detected the fewest non-word errors; the self-no study group detected the most. A two-way ANOVA with non-word error detection as the dependent measure revealed that the effect of essay condition was not significant ($F(1, 96) = 0.28$, $\eta_p^2 = 0.003$, $p = .599$), nor was the effect of study condition ($F(1, 96) = 1.51$, $\eta_p^2 = 0.016$, $p = .222$), or the Essay Condition × Study Condition interaction ($F(1, 96) = 1.51$, $\eta_p^2 = 0.016$, $p = .222$).

Content vs. function errors. Consistent with Experiment 1, participants detected significantly more content errors ($M = 91.13\%$, $SD = 11.28\%$) than function errors ($M = 82.00\%$, $SD = 14.20\%$), $t(99) = 6.40$, $d = 0.71$, $p < .001$.

Content errors (Fig. 10). A two-way ANOVA with content error detection as the dependent measure revealed

Table 4 Descriptive statistics for proofreading, eye tracking, and memory test in Experiment 2

Measure	Self-no study	Self-study	Other-no study	Other-study
Proofreading				
Total errors detected (%)	85.4 (14.0)	82.8 (12.9)	85.6 (08.5)	88.8 (07.5)
Word errors (%)	80.4 (18.8)	81.2 (16.7)	82.4 (11.3)	88.8 (11.3)
Non-word errors (%)	90.4 (11.4)	84.8 (13.6)	88.8 (10.5)	88.8 (09.7)
Content errors (%)	92.0 (13.9)	90.0 (12.0)	88.5 (10.8)	94.0 (07.3)
Function errors (%)	81.0 (16.8)	78.0 (17.3)	83.3 (11.0)	85.7 (09.8)
Eye tracking				
Number of fixations	808.6 (314.8)	707.0 (231.7)	804.7 (260.9)	857.2 (280.0)
Saccade amplitude (px)	44.8 (14.1)	45.5 (14.2)	42.5 (13.1)	40.3 (12.2)
Fixation duration (ms)	227.9 (28.3)	231.2 (27.0)	228.9 (21.6)	234.1 (32.9)
Time spent (s)	181.1 (65.4)	162.7 (53.0)	183.0 (57.4)	196.7 (56.2)
Memory test	9.6 (0.6)	9.8 (0.5)	9.6 (0.9)	9.2 (0.7)

Means are presented with *SD*s in parentheses. $N = 25$ per group for all measures except for the memory test, for which Self-No Study $N = 24$ and Other-No Study $N = 24$

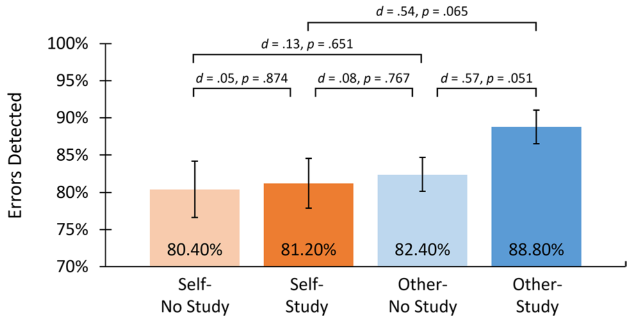

Fig. 8 Word error detection rates in Experiment 2

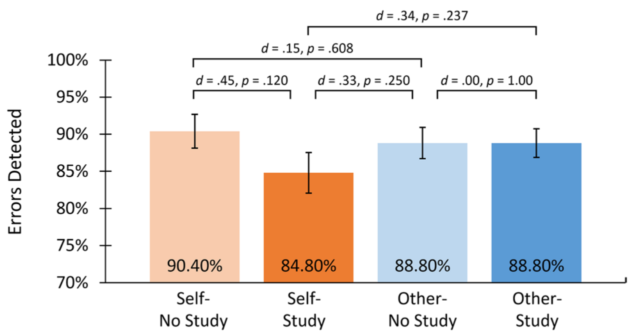

Fig. 9 Non-word error detection rates in Experiment 2

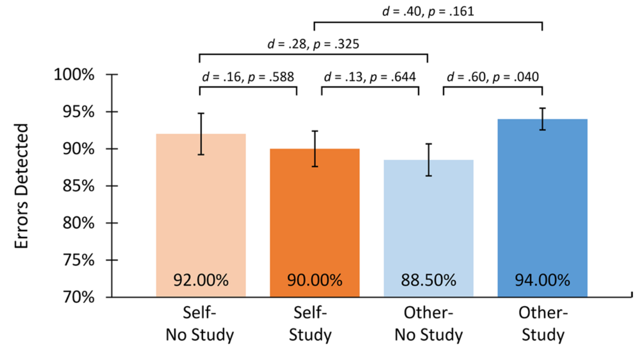

Fig. 10 Content error detection rates in Experiment 2

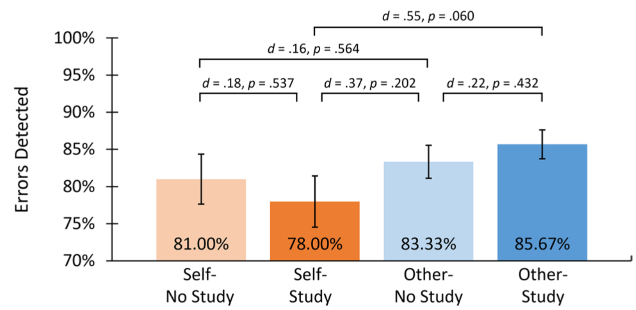

Fig. 11 Function error detection rates in Experiment 2

that the effect of essay condition was not significant (F(1, 96) = 0.01, $\eta_p^2 = 0.000$, $p = .912$), nor was the effect of study condition (F(1, 96) = 0.60, $\eta_p^2 = 0.006$, $p = .439$), or the Essay Condition × Study Condition interaction (F(1, 96) = 2.78, $\eta_p^2 = 0.028$, $p = .099$).

Function errors (Fig. 11). The self-study group detected the fewest function errors; the other-study group detected the most. A two-way ANOVA with function error detection as the dependent measure revealed that the effect of essay condition was not significant (F(1, 96) = 3.13, $\eta_p^2 =$ 0.032, $p = .080$), nor was the effect of study condition (F(1, 96) = 0.01, $\eta_p^2 = 0.000$, $p = .906$), or the interaction of Essay Condition × Study Condition (F(1, 96) = 0.89, $\eta_p^2 = 0.009$, $p = .348$).

Eye tracking. Although the self-study group made fewer fixations than the other-study group (t(48) = 2.07, $d = 0.59$, $p = .044$) and spent less time proofreading (t(48) = 2.20, $d = 0.62$, $p = .032$), the main effects of essay condition, study condition, and their interaction were not significant (all $ps > .12$; Table 4).

As in Session 1 of Experiment 1, number of fixations correlated with proofreading performance ($r = .22$, $p = .029$), as did saccade amplitude ($r = -.32$, $p = .001$; Table 5). Participants who made more fixations and shorter saccades detected more errors. The correlation between time spent proofreading and performance was non-significant ($r = .19$, $p = .058$).

Memory test. The self-study group ($M = 9.8$ out of 10) and self-no study group ($M = 9.6$ out of 10) had very high accuracy rates on the memory test, and the difference between them was not significant (t(47) = 0.86, $d = 0.25$, $p = .394$; Table 4). The other-study group performed significantly worse on the memory test than the self-no study group (t(47) = 2.30, $d = 0.66$, $p = .026$) and the self-study group (t(48) = 3.18, $d = 0.90$, $p = .003$). Thus, memory for one's own essay was stronger than memory for someone else's essay.

Individual differences. None of the individual-difference measures correlated significantly with proofreading (Table 5). As in Experiment 1, the correlation between reading comprehension and proofreading was non-significant ($r = .08$, $p = .427$). The two self-reported measures of academic achievement, ACT ($r = .16$, $p = .241$) and GPA ($r = .14$, $p = .266$), did not correlate significantly with proofreading, and neither measure of miserly cognitive processing, the cognitive reflection test ($r = .01$, $p = .909$) and need for cognition ($r = -.08$, $p = .463$), significantly predicted proofreading performance.

Table 5 Correlations between proofreading, eye tracking, cognitive ability, and personality in Experiment 2

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Proofreading performance	–													
2. Word errors	**.88**	–												
3. Non-word errors	**.77**	**.37**	–											
4. Content errors	**.71**	**.65**	**.51**	–										
5. Function errors	**.93**	**.81**	**.72**	**.39**	–									
6. Fixations	**.22**	.13	**.25**	.14	**.20**	–								
7. Saccade amplitude	**− .32**	**− .22**	**− .32**	**− .27**	**− .26**	**− .87**	–							
8. Fixation duration	− .05	− .02	− .07	− .01	− .06	**− .30**	**.28**	–						
9. Time spent proofreading	.19	.11	**.22**	.12	.18	**.91**	**− .79**	.09	–					
10. Memory test	− .15	− .17	− .07	− .13	− .14	.18	− .14	**− .26**	.09	–				
11. Reading comprehension	.08	.16	− .05	.16	.03	**− .27**	**.29**	.05	**− .27**	.11	–			
12. Cognitive reflection test	.01	− .03	.06	.02	.00	.00	.03	− .09	− .03	.10	**.36**	–		
13. Need for cognition	− .08	.07	**− .23**	.07	− .14	− .05	.14	.03	− .03	.13	**.40**	**.41**	–	
14. ACT	.16	.12	.14	.05	.18	.03	.01	.08	.14	− .09	**.62**	**.48**	**.39**	–
15. GPA	.14	.15	.06	.01	.16	− .09	.09	.19	− .03	.05	**.32**	.05	.14	**.36**

Bolded correlations are significant at $p < .05$. $N = 100$ for correlations between measures 1–9; for the remaining correlations, *Ns* range from 45 to 100

Discussion

Taken together, the results of the experiments are equivocal. They did not provide strong support for Daneman and Stainton's (1993) hypothesis that memory for self-generated text leads to an expectancy-driven processing style that makes proofreading more difficult. Even though Experiment 1 was a near replication of Daneman and Stainton's (1993) procedure, the results were in the reverse direction from the one predicted: students who proofread their own essays immediately after writing them detected 5.3% ($p = .059$) *more* errors than students who proofread someone else's. This result, although non-significant, is more consistent with Pilotti and Chodorow (2009), who found that familiarity for self-generated text facilitated proofreading performance, than with Daneman and Stainton (1993), who found the reverse. The results also revealed no difference between participants who proofread the essay they had written one week prior compared to those who proofread another participant's essay.

Experiment 2 used a studying manipulation to test whether inducing even greater familiarity with self-generated text would lead to a decrement in proofreading performance. More precisely, we predicted that studying another student's essay would facilitate proofreading, whereas studying one's own essay would impair proofreading. Indeed, students who studied and proofread their own essay detected the fewest errors, whereas students who studied and proofread someone else's essay detected the most errors. Once again, though, group differences were not significant, likely due to insufficient statistical power. The power-analyses that informed data collection for these experiments were based on the effect size observed by Daneman and Stainton (1993), which may be an overestimate of the true effect. On balance, then, our results must be viewed as preliminary, as follow-up studies with larger samples are needed.

Nevertheless, students who studied and proofread their own essay appeared to engage in more top-down processing; they made significantly fewer fixations and spent less time proofreading than students who studied and proofread someone else's essay. Thus, familiarity from having generated the text oneself may have led to what Levy et al. (1986) called *fluent reading*, an increase in efficiency without appreciable loss in visual or semantic analysis. We also found that eye movements and, in particular, indicators of bottom-up processing were positively correlated with proofreading performance. Across groups and experiments, number of fixations correlated with proofreading (*r*s ranged from .22 to .27) as did time spent proofreading (*r*s ranged from .19 to .35). As we noted in the Introduction and discuss further in the Limitations section, although evidence has linked eye movement patterns to levels of processing (Hyönä & Niemi, 1990; Kaakinen & Hyönä, 2010; Schotter et al., 2014), alternative interpretations are possible.

What can be made of these analyses? Hypothesized indicators of bottom-up processing were positively correlated with proofreading performance, and familiarity with the text was associated with less bottom-up processing, at least as inferred through patterns of eye movements. Although this pattern is broadly consistent with our hypotheses, the results did not support the critical finding that proofreading is impaired by familiarity with self-generated text. Our work can inform future efforts to study the self-generation effect in proofreading by providing a more conservative estimate of the effect size. That is, future studies should attempt to induce overfamiliarity for text using stronger manipulations to maximize the magnitude of the effect, and should collect data on much larger samples to provide adequate statistical power to detect smaller effects.

Across groups and experiments, non-word errors (misspellings) were more readily detected than word errors (wrong words). This finding is consistent with prior work by Levy et al. (1986) and Schotter et al. (2014); in both cases, participants detected non-word errors more frequently than word errors. Moreover, this finding has practical significance because word errors are also less likely to be detected by "spell check" than non-word errors. We also found that content errors were more readily detected than function errors. That said, function errors were significantly shorter ($M_{\text{characters}} = 3.8$) than content errors ($M_{\text{characters}} = 4.7$) ($t(106) = 4.11$, $d = 0.79$, $p < .001$). This could drive the difference in detection rates because shorter words are less likely to be fixated (Just & Carpenter, 1980; Rayner, 1998); mistakes embedded in function words might therefore be easier to overlook.

The experiments also assessed individual differences in proofreading performance. In Experiment 1, the correlation between proofreading performance across sessions was positive and significant ($r = .37$, $p = .009$); people who were strong proofreaders during Session 1 performed well during Session 2. Fluid intelligence correlated significantly with proofreading (rs ranged from .21 to .45), suggesting a link between novel problem-solving ability and performance in an error detection task and corroborating the results of Furnham (2010) and Furnham et al. (2006). It is possible that people with greater fluid intelligence are better able to reason to detect mistakes, or adapt to the novel demands of the experimental paradigm. By comparison, processing speed did not predict proofreading performance (rs ranged from − .03 to − .18), nor did working memory capacity (rs ranged from .03 to .04). In contrast to Daneman and Stainton (1993), reading comprehension scores did not significantly predict proofreading performance (rs ranged from .01 to .20). This latter result was unexpected. There was not substantial restriction of range in the reading comprehension scores which would artificially attenuate the correlation, and conceptually, those who are better able to make sense of passages would be expected to perform better at detecting nonsensical errors. Perhaps the skills required for reading comprehension and error detection overlap less than previously thought. Finally, miserly cognitive processing, as assessed by performance on the cognitive reflection test and need-for-cognition questionnaire, did not predict proofreading performance (rs = .01 and − .08, respectively).

Implications for the self-generation effect in context of proofreading

Only a few studies have examined whether students are worse at proofreading their own writing than the writing of others. The available evidence is mixed, but except for Daneman and Stainton (1993), it goes against the self-generation effect. Although our Experiment 1 was conducted as a near replication, the results trended in the reverse direction. Coupled with the results of Pilotti and Chodorow (2009), who found a significant facilitating effect of self-generated familiarity in a larger sample, it appears that the self-generation effect may be on tenuous ground.

What could explain these contradictory findings? There are a number of possibilities. Surely, small samples have led to substantial variability in effect sizes across studies. To their credit, Pilotti and Chodorow (2009) tested more than 100 participants and found no evidence for the self-generation effect. In our Experiment 2, we had 100 participants divided into four groups. Another possibility is that only extreme familiarity impairs proofreading, of the kind that is developed by working on a piece extensively over the course of days, weeks, or months. Our Experiment 2 provides tentative support for this possibility (we found a small, non-significant self-generation effect after attempting to induce overfamiliarity for self-generated text). Of course, testing this hypothesis in a 'real-world' context would be challenging, and a first step might be to develop better procedures to induce overfamiliarity for self-generated text in the laboratory. This may pose difficulty; in our Experiment 2, performance on a memory test for self-generated text was nearly perfect ($M = 9.6$ out of 10) and increased only marginally following a studying manipulation ($M = 9.8$ out of 10; a non-significant difference). Ceiling effects resulting from the use of recognition-based questionnaires may partly explain this finding. Other methods, such as asking participants to freely recall what they wrote, may lead to greater variance in performance, and more clearly indicate an effect of studying on familiarity with self-generated text.

In future work, it would also be worthwhile to assess participants' memory for self-generated text as a function of the length of time between writing and proofreading. Daneman and Stainton (1993) suggested that after two weeks, participants' overfamiliarity with their essays decreased, reducing their use of expectancy-driven processing and improving their proofreading performance. We included a one-week delay in our Experiment 1; however, we did not find that it improved proofreading performance. That said, while it is generally true that memory fades with the passage of time (Ebbinghaus, 1885), a memory test in Experiment 1 would have provided direct evidence to test this claim within the context of self-generated text. Specifically, one could test whether changes in proofreading performance and eye-tracking measures correlate with changes in memory test performance.

Finally, the widespread use of computers has changed the nature of proofreading. Daneman and Stainton (1993) gave students a printout of the essay they had written and instructed them to mark any errors that they noticed. So far as we know, every study thereafter has used a computerized proofreading task. Notwithstanding the difference in method across studies, there are also implications for ecological validity because modern spell-checkers can detect non-word errors. They also are increasingly able to detect wrong words embedded in sentences. Although writers should always be careful to say what they mean, how this affects proofreading as a step in the writing process may change as technology continues to advance.

Limitations

One limitation of the present work is that although adding errors to students' essays afforded experimental control, it also threatened external validity. That is, it is unusual to be asked to proofread an essay after being told that errors have been added to it. However, if we left students' errors in their essays instead of adding our own, the number and kind of mistakes would differ across essays, and we would not know whether students were aware of their mistakes without asking them. Research on the self-generation effect has followed the same procedure we used here; errors were added by the experimenter to participants' essays (e.g., Daneman & Stainton, 1993; Pilotti & Chodorow, 2009). Despite the threat to external validity, a real-world analogue of this approach might be receiving copyedits from an editorial office; sometimes, a well-intentioned editor makes unwanted changes to a manuscript, which the author must then detect and correct.

Another limitation of our work is the difficulty of directly observing top-down and bottom-up processing during reading and proofreading. Throughout this report, we described bottom-up processing as reflecting close textual analysis and evinced by eye movement patterns such as more fixations, longer fixations, and shorter saccades. By comparison, we argued that top-down processing would be characterized by larger saccades, with participants "filling in the blanks" using expectancy-driven processing. While studies have shown that when participants know more about a passage (facilitating top-down processing), they make fewer fixations and larger saccades (Hyönä & Niemi, 1990; Raney et al., 2000), there is not necessarily a one-to-one correspondence between eye movements during reading and levels of processing—other factors could play a role. Although this issue is ubiquitous in cognitive psychology (i.e., unobservable cognitive processes are inferred through observable behaviors), it is worth acknowledging that the link between eye movements and levels of processing reflects psychological theory, and alternative interpretations are possible.

Conclusion

Two eye-tracking experiments were conducted on the self-generation effect in proofreading. In Experiment 1, the results were contrary to prediction and in Experiment 2, consistent with prediction but not significant. Given the theoretical importance of the self-generation effect, we believe that its further study will be worthwhile, especially with manipulations designed to induce extreme familiarity for self-generated text and with larger samples with adequate statistical power to detect small effects.

Funding This research was funded by the Future Academic Scholars in Teaching Fellowship awarded to Alexander P. Burgoyne.

Availability of data Data and materials will be made available upon request.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading, 18*(5), 325–346.

Barach, E., Gloskey, L., & Sheridan, H. (2021). Satisfaction-of-Search (SOS) impacts multiple-target searches during proofreading: Evidence from eye movements. *Visual Cognition, 29*, 510–518.

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-denny reading test forms G&H*. Houghton Mifflin Harcourt.

Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (2019). Is working memory capacity a causal factor in fluid intelligence? *Psychonomic Bulletin & Review, 26*, 1333–1339. https://doi.org/10.3758/s13423-019-01606-9

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Carrell, P. L., & Eisterhold, J. C. (1983). Schema theory and ESL reading pedagogy. *TESOL Quarterly, 17*, 553–573. https://doi.org/10.2307/3586613

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466.

Daneman, M., & Stainton, M. (1993). The generation effect in reading and proofreading. *Reading and Writing, 5*, 297–313. https://doi.org/10.1007/BF01027393

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. Dover.

Ekstrom, R. B., French, J. W., Harmon, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests*. Educational Testing Service.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Furnham, A. (2010). Proofreading as an index of crystallised intelligence. *Educational Psychology, 30*, 735–754. https://doi.org/10.1080/01443410.2010.506005

Furnham, A., Rawles, R., & Iqbal, S. (2006). Personality, intelligence and proof-reading. *Personality and Individual Differences, 41*, 1457–1467. https://doi.org/10.1016/j.paid.2006.04.025

Gallagher, D. J., & Hall, S. J. (1992). Proof-reading performance of extraverts and neurotics under punishment and reward conditions. *Personality and Individual Differences, 13*(2), 229–235. https://doi.org/10.1016/0191-8869(92)90047-S

Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, *73*, 259–280. https://psycnet.apa.org/doi/10.1016/0001-6918(90)90026-C

Inhoff, A. W., Topolski, R., Vitu, F., & O'Regan, J. K. (1993). Attention demands during reading and the occurrence of brief (express) fixations. *Perception & Psychophysics, 54*, 814–823. https://doi.org/10.3758/BF03211806

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. Greenwood Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://psycnet.apa.org/doi/10.1037/0033-295X.87.4.329

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1561–1566. https://doi.org/10.1037/a0020693

Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes, 39*, 125–128. https://doi.org/10.1080/0163853X.2005.9651676

Levy, B. A. (1983). Proofreading familiar text: Constraints on visual processing. *Memory & Cognition, 11*, 1–12. https://doi.org/10.3758/BF03197655

Levy, B. A., Newell, S., Snyder, J., & Timmins, K. (1986). Processing changes across reading encounters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 467–478. https://psycnet.apa.org/doi/10.1037/0278-7393.12.4.467

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods, 47*, 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

Paterson, K. B., McGowan, V. A., Warrington, K. L., Li, L., Li, S., Xie, F., Chang, M., Zhao, S., Pagán, A., White, S., & Wang, J. (2020). Effects of normative aging on eye movements during reading. *Vision, 4*(1), 7. https://doi.org/10.3390/vision4010007

Pilotti, M., & Chodorow, M. (2009). Error detection/correction in collaborative writing. *Reading and Writing, 22*, 245–260. https://doi.org/10.1007/s11145-007-9110-x

Raney, G. E., Therriault, D. J., & Minkoff, S. R. B. (2000). Repetition effects from paraphrased text: Evidence for an integrated representation model of text representation. *Discourse Processes, 29*(1), 61–81. https://doi.org/10.1207/S15326950dp2901_4

Raven, J., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://psycnet.apa.org/doi/10.1037/0033-2909.124.3.372

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–776. https://psycnet.apa.org/doi/10.1037/0012-1649.27.5.763

Schmauder, A. R., Morris, R. K., & Poynor, D. V. (2000). Lexical processing and text integration of function and content words: Evidence from priming and eye fixations. *Memory & Cognition, 28*, 1098–1108. https://doi.org/10.3758/BF03211811

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition, 131*, 1–27. https://doi.org/10.1016/j.cognition.2013.11.018

SR Research Ltd. (2010). Eyelink 1000 users' manual, version 1.5. Mississauga

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729