Fortune Favors the Bold (and the Italicized): Effects of Disfluency on Educational Outcomes

Daniel M. Oppenheimer (doppenhe@princeton.edu)
Princeton University, Department of Psychology
Green Hall, Princeton, NJ 08540 USA

Connor Diemand Yauman (cdiemand@princeton.edu)
Princeton University, Department of Psychology
Green Hall, Princeton, NJ 08540 USA

Erikka B. Vaughan (ebvaugha@umail.iu.edu)
Indiana University, Department of Psychology
1101 E. 10th Street, Bloomington, IN 47405

Abstract

Research has shown that disfluency – the metacognitive experience of difficulty associated with a cognitive task – engenders deeper processing. Since deeper processing typically leads to better retention, this paper examined whether decreasing perceptual fluency of educational materials would improve retention. Study 1 found that harder to read fonts led to increased retention in a controlled laboratory setting. Study 2 extended this finding to real-world classroom environments. It appears as though perceptual disfluency can function as a desirable difficulty in education. Implications and caveats are discussed.

Introduction

It seems logical that to effectively communicate an idea, one should present it in a manner which is clear and easy to follow. Educators follow this principle when designing textbooks—the order, wording, and formatting is designed to help students read the information with minimal effort. Indeed, there is evidence to support the notion that students benefit from decreased cognitive demands when learning new concepts (Sweller and Chandler, 1994).

While it is commonly accepted that reducing extraneous cognitive load is beneficial to student learning, there is some research that seems to suggest there are exceptions to this rule. In fact, research shows that in certain instances, it may be beneficial to *increase* extraneous cognitive load (e.g. Bjork 1994). These aptly named "desirable difficulties" create additional cognitive burdens but nonetheless improve learning.

For example, in one experimental paradigm (Hirshman & Bjork, 1988), participants are asked to remember pairs of words, such as "bread : butter." Hirshman and Bjork found that requiring subjects to mentally generate missing letters in a word pair, such as "bread : b_tt_r," leads to improved recall performance over participants who read the word pair without any missing letters. Bjork extended this strategy to realistic educational settings, finding that students who complete simple fill-in-the-blank sentences are better able to retain information than students who read the same sentences with the key words filled in and underlined for them (Richland, Bjork, Finley, & Linn, 2005).

It seems counterintuitive that imposing unnecessary strain on students' limited cognitive capacity would actually improve performance, yet desirable difficulties seem to exploit nuances in our cognitive systems. Importantly, these instructional techniques *appear* sub-optimal. Without conscious recognition and implementation on behalf of cognitive psychologists and educators, it is likely that these techniques would not even be considered for use.

It is important to explore such techniques and seek out new methods of presentation that better reflect or utilize the way we process information. One such technique may come from explorations on the metacognitive experience of fluency—the subjective feeling of ease or difficulty which is associated with almost any mental task (Alter & Oppenheimer, 2009). For instance, a blurry photograph is disfluent because it is difficult to discern, a whisper is disfluent because it is difficult to hear, and a foreign word may be disfluent because it is difficult to pronounce. Fluency has been shown to influence our judgments in a variety of ways, including our judgments of truth, confidence, intelligence, or familiarity (for a review, see Alter & Oppenheimer, 2009). Importantly, recent studies have begun to explore how fluency influences cognitive processing in ways that might yield positive educational outcomes.

Recent work in fluency has demonstrated that when a problem is disfluent, people adopt a more deliberate processing strategy (Alter, Oppenheimer, Epley, & Eyre, 2007). In one experiment, participants were asked to read logical syllogisms and indicate whether they were true or false. Participants who read the syllogisms in a difficult to read (i.e. disfluent) font performed significantly better on the task than those who read the syllogisms in a clear, easy to read font. The authors replicated this result in three distinct cognitive domains. In this way, disfluency may be categorized as a desirable difficulty and can be used to improve student learning by encouraging them to select more accurate problem solving strategies.

Contemporary educational reform measures strive to create learning environments that encourage students to engage deeply with course content because of the numerous forms of evidence suggesting that deep processing increases learning. Most importantly, deeper processing facilitates later recall. For example, participants who are asked whether words appear in capital or lower case letters (low level of processing) do worse at later recall than participants who construct a rhyme for the words (moderate processing) or are asked to define the words (deep processing) (Craik & Tulving, 1975). Therefore, if disfluency facilitates deeper processing, then there is reason to expect that disfluent educational materials will lead to improved retention in the classroom.

The simplest and most standard fluency manipulation is a font manipulation. Information presented in an easy to read font is more fluent than *information presented in a difficult to read font*. The beauty of this manipulation is that it is so easy and cost effective to implement. To the extent that disfluency yields better learning outcomes, the intervention could be implemented on a wide scale with limited logistical or financial challenges.

The purpose of the present research is to empirically examine whether fluency can operate as a desirable difficulty to improve retention in classroom environments.

Study 1

First we aimed to show that disfluency led to better retention in a highly controlled laboratory environment. Twenty eight participants were recruited through the Princeton University paid subject pool and compensated $8 for their time. Participants' ages ranged from 18-33. Participants were given 90 seconds to learn about three species of aliens. Each alien species had seven features, for a total of 21 features that needed to be learned (see Figure 1 or examples of the features to be learned). This task was meant to approximate taxonomic learning that might occur in a biology classroom; fictional alien species were used so that participants had no prior knowledge that might contaminate results.

In the disfluent condition, the stimuli were presented in either 12 point Comic Sans MS 75% greyscale (see Figure 1a) or 12 point *Bodoni MT* 75% grayscale font. In the fluent condition, the stimuli were presented in 16-point Arial 100% black font (See Figure 1b). A between-subjects design was used, such that each participant was only exposed to one font. As is evident from the examples below, while the disfluent text is obviously harder to read than the fluent text (when they are presented side by side) in a between subject design reader's in the disfluent condition were unlikely to even consciously notice the added difficulty the disfluent text engendered.

The pangerish

- *Ten feet tall*
- *Eats green, leafy vegetables*
- *Has blue eyes*

The norgletti

- Two feet tall
- Eats flower petals and pollen
- Has brown eyes

Figure 1: Example stimuli from Study 1. The top panel shows the disfluent font, and the bottom panel shows the fluent font.

After studying the material for 90 seconds, participants were distracted for 15 minutes with unrelated tasks. Participants' memory for the material was then tested. For each participant, seven of the features were randomly asked about. For example "how tall is the pangerish?" or "what color eyes does the norgletti have?"

One outlier was eliminated from consideration for being more than 3 standard deviations from the mean. Participants in the fluent condition were accurate 72.8% of the time. Meanwhile, participants in the disfluent condition successfully remembered the information 86.5% of the time. This difference was statistically significant ($t(26) = 2.3$, $p < .05$). There were no differences in retention between the different disfluent fonts (Comic Sans vs. Bodoni), suggesting that it was not the specific font that led to the difference, but rather the disfluency. In sum, after a 15-minute delay, participants in the recalled nearly 15% more information when the material was presented disfluently than fluently. Moreover, as learning time was constrained, this cannot be due to longer study times for the disfluent materials, which suggests that instead more effective learning strategies were adopted.

While this provides strong preliminary evidence that fluency could be a desirable difficulty in education, there are several reasons why we might be concerned about its generalizability to actual classroom environments. First, the materials we used, while tightly controlled, were not the sorts of materials that would be used in real classroom settings. Different types of materials might elicit different effects. Second, while the effects in Study 1 persisted for 15 minutes, the time between learning and testing is typically much longer in the real world.

Further, while paid laboratory participants may be willing to persist in the face of challenging fonts for 90 seconds, added difficulty may undermine motivation for actual students. Students may just give up, rather than deeply processing the material – particularly as the semester progresses and stress levels rise. Therefore, we ran a large field study to determine whether these results would persist outside of the lab.

Study 2

222 high school students (ages 15-18) from a public school in Chesterland, Ohio participated in the study. This school accommodates approximately 930 students from grades 9-12 and reported a 98.6% graduation rate in 2008. The school's grades 9-12 are taught by 54 teachers.

Classes were selected for this research using the following criteria: the same teacher must have been teaching at least two classes of the same subject and difficulty level with the same supplementary learning material (PowerPoint presentations or handouts). Six classes met these criteria and agreed to participate. These classes were AP English, Honors English, Honors Physics, Regular Physics, Honors US History, and Honors Chemistry.

The different sections of each class were randomly assigned to either a disfluent or control category. Teachers were instructed to send all relevant supplementary learning materials to the experimenters prior to distributing them to students. At no point did the experimenters ever have face-to-face contact with the students or teachers; editing was done by proxy in Princeton, New Jersey. The fonts of the learning material in the disfluent category were either changed to **Haettenschweiler,** *Monotype Corsiva,* or *Comic Sans Italicized* or copied disfluently (by moving the paper up and down during copying) when electronic documents were unavailable. In the control category, no edits were made to the materials before returning them to teachers. The font size of the supplementary material was not changed unless the original size when converted to disfluent font made the font illegible, in which case the font size was increased until it was readable. One teacher refused to administer Haettenschweiler and so that class was changed to Comic Sans Italicized.

No other changes were made to the students' learning environments, materials, curricula, or to the teachers' classroom routine. To determine the effects of disfluency, the results of the normal assessment tests for the class were collected and analyzed.

The z-scores of the students' test performance were used as a common metric to compare students across different courses. As shown in table 1, average z-scores of the students were higher in the disfluent condition than in the control.

An independent samples t-test of the average z-scores revealed a significant improvement of the students' test scores in the disfluent condition ($t(220) = 3.38$, $p < .001$): students in the disfluent condition scored higher on their tests (M= .164, SD = .1.03) than those in the control (M = -.295, SD = 1.05). There were no reliable differences between the different disfluent fonts. That is, it was not the specific of the font that mattered, but rather the fact that it was disfluent.

	Control	Disfluent
AP English	-.058	.135
Honors English	-.175	.131
Physics Honors	-.251	.215
Physics Normal	-1.13	.42
History	-.177	.112
Chemistry	.023	-.017
Total	-.295	.164

Table 1: Average z-score for fluent and disfluent supplementary materials across the 5 usable classrooms. Note that the z-scores do not sum to 0 across conditions because of unequal sample sizes by condition.

The effects of different kinds of disfluent material were examined using a two-level ANOVA to compare the effects of disfluent worksheets and PowerPoint presentations. This test revealed that the PowerPoint presentations were significantly more effective than the documents in improving student performance when presented in a disfluent format ($F(1, 184) = 9.38$, $p < .01$). However it is difficult to read too deeply into this latter finding, as the only classes that used powerpoint materials were the physics classes. As such, we cannot know if the difference was due to the type of material that was being studied, or the manner in which it was presented. Nonetheless, the difference highlights possible future avenues of exploration.

Discussion

In two studies we showed that making the text disfluent by using a hard to read font improved learning. In Study 1, participants recalled 14% more material when the material was initially presented in a disfluent font. In Study 2, students performed better on exams in actual classrooms the fonts of the supplementary materials were harder to read. This occurred for both science and non-science courses, and for different difficulty levels (AP, honors, and regular). This provides strong preliminary evidence that disfluency can indeed function as a desirable difficulty in educational settings.

There are, however, some important caveats that need to be considered in relation to these findings. First, while a small amount of disfluency was able to improve performance, at some level disfluency will necessarily impair functioning. After all, if the font is impossible to read, then the information cannot be encoded, let alone retained. It is unclear from these studies what the optimal level of disfluency is, nor the relative detriment that being overly disfluent might engender.

Secondly, there is the issue of adaptation. One reason that disfluent text might lead to better retention is that it serves as an alarm signal that this material is challenging and merits extra consideration (c.f. Alter et al., 2007). To the extent that students become used to disfluency, they might no longer adopt deeper processing strategies when exposed to hard to read font. Study 2 was limited to a single semester's worth of materials for logistical reasons. It is unclear whether these effects would persist over longer periods of time.

Third, a large literature has demonstrated that disfluent materials are liked less than fluent materials (for a review see Alter and Oppenheimer, 2009). It may be that while students retained the information better under disfluency, they also liked the material less. This could mean that they are less likely to pursue further studies in the topic (e.g. in college) or are otherwise demotivated in subsequent educational situations. Of course, it is also possible that the increased effort necessary to engage with this material will create cognitive dissonance, which will cause them to like the material more (c.f. Cooper, 2007). This is an empirical question, that will require additional research to resolve.

Fourth, it is quite possible that there are moderators for this effect that these initial studies did not detect. Other forms of desirable difficulties have been shown to be moderated by factors such as the nature of the materials (McDaniel et al., 2000) the nature of the testing (Thomas & McDaniel, 2007), and the abilities of the learner (Macnamara et al., 1996). One could imagine that less motivated students from a less successful school might be more inclined to give up on the material rather than persist and encode it more deeply. Future investigation should look into these issues.

Despite these potential drawbacks, disfluency is a promising form of desirable difficulty because it requires no retraining of teachers, no restructuring of curricula, and can be implemented with minimal cost. Given the results of the present studies, it seems worthwhile to investigate disfluency as an educational intervention further.

Acknowledgments

Thanks to Dr. Beth Yauman, a Princeton Undergraduate research fellowship and the PSURE research fellowship for supporting the second and third authors respectively. Thanks to Michelle Ellefson, Alan Castel , Adam Alter, Anuj Shah, Jeff Zemla, and the Opplab for advice and support.

References

Alter, A.L., & Oppenheimer, D.M. (2009). Uniting the tribes of fluency to form a metacognitive nation. Personality and Social Psychology Review. *Personality and Social Psychology Review. 13(3),* 219-235.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *JEP: General*, *136*, 569-576.

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.

Cooper, J. (2007). *Cognitive dissonance: 50 years of a classic theory*. London: Sage publications.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. J*EP: General*, *104*, 268-294

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 484-494.

McDaniel, M.A., DeLosh, E.L., & Merritt, P.S. (2000). Order information and retrieval distinctiveness: Recall of common versus bizarre material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1045-1056.

McNamara, D. S., Kintsch, E., Butler-Songer, N., and Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction , 14* (1), 1-43.

ODE - Home. Web. 21 Jan. 2010. <http://www.ode.state.oh.us>.

Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: generation and interleaving effects. In B. G. Bara, L. Barsalou and M. Bucciarelli (Eds.) *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*, 185-233.

Thomas, A.K., & McDaniel, M.A. (2007). Metacomprehension for educationally relevant materials: Dramatic effects of encoding-retrieval interactions. *Psychonomic Bulletin and Review, 14,*212-218.