Cognition xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

# Fortune favors the Bold (and the Italicized): Effects of disfluency on educational outcomes

Connor Diemand-Yauman [a], Daniel M. Oppenheimer [a,*], Erikka B. Vaughan [b]

[a] Princeton University, Department of Psychology, United States
[b] Indiana University, Psychological & Brain Sciences, United States

ARTICLE INFO

*Article history:*
Received 5 May 2010
Revised 3 September 2010
Accepted 8 September 2010
Available online xxxx

*Keywords:*
Fluency
Education
Desirable difficulties

ABSTRACT

Previous research has shown that disfluency – the subjective experience of difficulty associated with cognitive operations – leads to deeper processing. Two studies explore the extent to which this deeper processing engendered by disfluency interventions can lead to improved memory performance. Study 1 found that information in hard-to-read fonts was better remembered than easier to read information in a controlled laboratory setting. Study 2 extended this finding to high school classrooms. The results suggest that superficial changes to learning materials could yield significant improvements in educational outcomes.



## 1. Introduction

Many educators believe that their ability to teach effectively relies on instinct and experience (Book, Byers, & Freeman, 1983). However, research has shown that instinct can be deceiving and lead to educational strategies that are detrimental to learners (Bjork, 1994). For example, students tend to gauge the relative success of a learning session based on the ease of encoding information rather than subsequent performance, and instructors may not have access to information on student's long-term retention and thus may also evaluate learning based on encoding (Bjork, 1994). Similarly, many education researchers and practitioners believe that reducing extraneous cognitive load is always beneficial for the learner (Sweller & Chandler, 1994). In other words, if a student has a relatively easy time learning a new lesson or concept, both the student and instructor are likely to label the session as successful even if the student is unable to retrieve the information at a later time.

However, in some cases making material harder to learn can improve long-term learning and retention (Bjork, 1994). More cognitive engagement leads to deeper processing, which facilitates encoding and subsequently better retrieval (Craik & Tulving, 1975). Aptly named "desirable difficulties" capitalize on this by creating additional cognitive burdens that improve learning.

For example, in one experimental paradigm researchers increased depth of processing by requiring the learner to generate rather than passively consume information (Richland, Bjork, Finley, & Linn, 2005). Hirshman and Bjork (1988) found that requiring participants to generate letters in a word pair (e.g. "salt:p_pp_r") during memorization resulted in a higher retention rate of the word pairs than when the pairs were presented in their entirety (e.g. "salt:pepper"). This was later extended this to real-world classroom environments and shown to yield similarly positive effects (Richland et al., 2005). It is worth noting that it is not the difficulty, per se, that leads to improvements in learning but rather the fact that the intervention engages processes that support learning. Thus, the benefits of generation can (under the right circumstances) overcome the drawbacks of the increased difficulty. Not all difficulties are desirable, and presumably interventions that engage more elaborative processes without also increasing

* Corresponding author. Address: Princeton University, Department of Psychology, Green Hall, Princeton, NJ 08540, United States. Tel.: +1 609 258 7465.

*E-mail address:* doppenhe@princeton.edu (D.M. Oppenheimer).


doi:10.1016/j.cognition.2010.09.012

Please cite this article in press as: Diemand-Yauman, C., et al. Fortune favors the Bold (and the Italicized): Effects of disfluency on educational outcomes. *Cognition* (2010), doi:10.1016/j.cognition.2010.09.012

difficulty would be even more effective at improving educational outcomes.

While researchers have identified a variety of desirable difficulties such as interleaving (Richland et al., 2005) and generation (Hirshman & Bjork, 1988; Slamecka & Graf, 1978), these methods simultaneously manipulate both the objective difficulty and subjective difficulty of encoding the material. This distinction is important, as disfluency – the subjective, metacognitive experience of difficulty associated with cognitive tasks – has been shown to impact cognitive processing independent of the objective cognitive difficulty (Alter & Oppenheimer, 2009; Oppenheimer, 2008).

There is strong theoretical justification to believe that disfluency could lead to improved retention and classroom performance. Disfluency has been shown to lead people to process information more deeply (Alter, Oppenheimer, Epley, & Eyre, 2007), more abstractly (Alter & Oppenheimer, 2008), more carefully (Song & Schwarz, 2008), and yield better comprehension (Corley, MacGregor, & Donaldson, 2007), all of which are critical to effective learning.

Importantly, disfluency can function as a cue that one may not have mastery over material (for a review, see Alter and Oppenheimer (2009)). For example, studies have shown that fluency is highly related to people's confidence in their ability to later remember new information (e.g. Castel, McCabe, & Roediger, 2007). To the extent that a person is less confident in how well they have learned the material, they are likely to engage in more effortful and elaborative processing styles (Alter et al., 2007).

For example, Alter and his colleagues presented participants with logical syllogisms in either an easy- or difficult-to-read font. Participants were significantly less confident in their ability to solve the problems when the font was hard-to-read, however they were in reality significantly more successful. Alter et al. (2007) subsequently showed that when material was disfluent participants were less likely to use heuristics, and tended to rely on more systematic and elaborative reasoning strategies. In this way, disfluency might indirectly improve retention and transfer by leading people to engage in deeper processing of the information (c.f. Oppenheimer, 2008).[1]

Disfluency can be produced merely by *adopting fonts that are slightly more difficult to read* (Alter & Oppenheimer, 2009). While other forms of desirable difficulties require significant curriculum reform to implement, disfluency manipulations could be adopted cheaply, easily, and without imposing on teachers.

To test whether disfluency based interventions could improve retention, participants learned fictional biological taxonomies that were presented in either easy or challenging fonts.

[1] Craik and Lockhart's (1972) demonstrations of depth of processing asked people to make judgments about surface features such as font, and showed that led to shallow processing. In contrast, while we are presenting materials in different fonts, we are asking participants to learn the material. We do not argue that participants attend to the font (which would be shallow processing) but rather that the font creates a metacognitive experience of disfluency which leads them to engage in more elaborative encoding strategies (which is deeper processing).

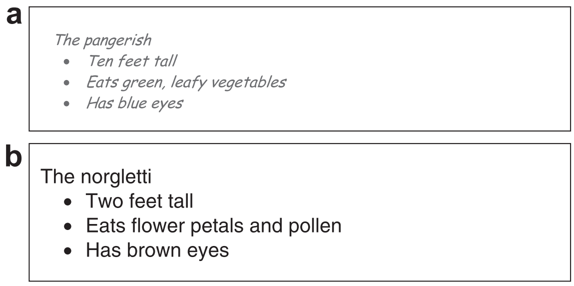


**Fig. 1.** Example stimuli from Study 1. The top panel shows an example disfluent font, and the bottom panel shows the fluent font.

## 2. Study 1

### 2.1. Participants

Twenty-eight participants recruited through the Princeton University paid subject pool took part in this study in exchange for $12. Participants' ages ranged from 18 to 40.

### 2.2. Materials, procedure, and design

Participants were asked to learn about three species of aliens, each of which had seven features, for a total of 21 features that needed to be learned (see Fig. 1 for examples of the features to be learned). This task was meant to parallel taxonomic learning in a biology classroom; alien species were used in place of actual species to ensure participants had no prior knowledge of the domain.

In the disfluent condition, material was presented in 12-point Comic Sans MS 60% grayscale font (see Fig. 1a) or 12-point *Bodoni MT* 60% grayscale. In the fluent condition, material was presented in 16-point Arial pure black font (see Fig. 1b). As is evident from the examples, the disfluency manipulation is quite subtle. While there is no question that the disfluent text feels harder to read than the fluent text when they are presented side by side, in the absence of a fluent sample to contrast against, it is unlikely that a reader would even be consciously aware of the added difficulty that the disfluent text engenders. A between-subjects design was used, such that each participant was only exposed to one font.

Participants were given 90 s to memorize the information in the lists. They were then distracted for approximately 15 min with unrelated tasks. Finally, participants' memory for the material was tested. For each participant, seven of the features were randomly sampled and participants were asked questions about those features (e.g. "what is the diet of the pangerish?" or "what color eyes does the norgletti have?").

## 3. Results and discussion

One participant was eliminated from analysis as an outlier for being more than three standard deviations from the mean. On average, participants in the fluent condition successfully answered 72.8% of the questions. Meanwhile, participants in the disfluent conditions were successful on

Please cite this article in press as: Diemand-Yauman, C., et al. Fortune favors the **Bold** (*and the Italicized*): Effects of disfluency on educational outcomes. *Cognition* (2010), doi:10.1016/j.cognition.2010.09.012

average 86.5% of the time. This difference was statistically significant ($t(26) = 2.3$, $p < .05$). In sum, after a 15-min delay, participants in the disfluent condition recalled 14 percentage points more information than those in the fluent condition. There were no reliable differences in retention between participants exposed to the different disfluent fonts. That is, the type of font that created the disfluency did not matter, merely that the font was disfluent.

While this result is encouraging, there are a number of reasons why this result might not generalize to actual classroom environments. First, while the effects persisted for 15 min, the time between learning and testing is typically much longer in school settings. Moreover, there are a number of other substantive differences between the lab and actual classrooms, including the nature of materials, the learning strategies adopted, and the presence of distractions in the environment, which could impede generalizability.

Another concern is that because disfluent reading is, by definition, perceived as more difficult, less motivated students may become frustrated. While paid laboratory participants are willing to persist in the face of challenging fonts for 90 s, the increase in perceived difficulty may provide motivational barriers for actual students. Moreover, a negative affective response could influence attitudes towards the material being learned and undermine motivation to engage in future study of the topic. Indeed, there is ample evidence that disfluency typically leads to reduced liking (Reber, Winkielman, & Schwarz, 1998).

To determine whether effects persisted in the real world and determine if there were negative affective consequences from the intervention, a second study was run in actual high school classrooms.

## 4. Study 2

### 4.1. Participants

Two hundred and twenty-two 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 (90% continue onto further education) and 95% attendance rate in 2008. Ninety-eight percent of students self-identify as White.

For a high school class to be a candidate for this research, 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 different science and non-science classes met these criteria. These classes were AP English, Honors English, Honors Physics, Regular Physics, Honors US History, and Honors Chemistry.

### 4.2. Materials, procedure, and design

Teachers were instructed to send all relevant supplementary learning materials to the experimenters in advance of their distribution. The experimenters received and manipulated two types of learning materials from teachers: worksheets and PowerPoint slides. However, PowerPoint slides were only available in physics classrooms. The actual content of the material was not altered in any way. At no point did the experimenters have face-to-face contact with the students or teachers; the editing of the materials was done by proxy in Princeton, New Jersey.

The different sections of each class were randomly assigned to a disfluent or control category. The fonts of the learning material in the disfluent condition were either changed to **Haettenschweiler**, *Monotype Corsiva*, *Comic Sans Italicized*, or were copied disfluently (by moving the paper up and down during copying) when electronic documents were unavailable. In the control condition, the learning materials were unedited. The font size of the supplementary material was not changed unless the size coupled with the disfluent font made the text illegible as reported by the teacher or experimenters, in which case the font size was adjusted to allow legibility. In one case a teacher refused to administer the Haettenschweiler font because he believed it was too hard-to-read. This class was switched to Comic Sans Italicized.

In an effort to prevent the well-documented effects of self-fulfilling prophecies, (Brophy, 1983; Haynes & Johnson, 1983) teachers were blind to the hypothesis and told only that the experiment focused on the effects of presenting reading in different fonts. It is likely that they would intuitively predict that the degraded font would cause students to perform more poorly, thus making the hypothesis more conservative by pitting it against expectancy effects.

No other changes were made to the students' learning environments or to the teachers' classroom routine. The length of the study in each individual classroom varied depending on the length of the teacher's lesson plan, ranging from a week and a half for history to nearly a month for physics. To determine the effects of disfluency, the results of the normal assessment tests for the class were collected and analyzed.

After the units were completed and exams were taken, a four-question survey was administered to test whether disfluency affected motivational factors. Students were asked to rate their responses to the following questions on a scale of 1–5: "How difficult do you find the material in this class?" (1 = 'very easy', 5 = 'very difficult'), "How do you feel about the material covered in class?" (1 = 'I like it very little', 5 = 'I like it very much'), "How frequently do you feel confused or lost during class?" (1 = 'never', 5 = 'all the time'), and "How likely are you to take classes on this material in college?" (1 = 'very unlikely', 5 = 'very likely').

## 5. Results

Student test performance was converted to $Z$-scores so as to provide a common metric to compare students across different courses. $Z$-scores were calculated across different sections of the same course, but not across courses. Students in the disfluent condition scored higher on classroom assessments ($M = .164$, $SD = 1.03$) than those in the control ($M = -.295$, $SD = 1.03$). An independent samples $t$-test revealed that this trend was statistically significant ($t(220) = 3.38$, $p < .001$, Cohen's $d = 0.45$, see Table 1 for a

Please cite this article in press as: Diemand-Yauman, C., et al. Fortune favors the **Bold** (*and the Italicized*): Effects of disfluency on educational outcomes. *Cognition* (2010), doi:10.1016/j.cognition.2010.09.012

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

| Course | Control | Disfluent |
|---|---|---|
| English AP | −.058 | .135 |
| English Honors | −.175 | .131 |
| Physics Honors | −.251 | .215 |
| Physics Regular | −1.13 | .421 |
| Chemistry | .023 | −.017 |
| History | −.177 | .097 |
| Average | −.295 | .164 |

class by class breakdown). There were no reliable differences between the particular fonts used (**Haettenschweiler**, *Monotype Corsiva*, *Comic Sans Italicized*, all $p > .1$).

In the follow-up survey assessing the students' feelings towards the material, an independent samples *t*-test of the average *Z*-scores revealed that there was no significant difference between the disfluent and fluent samples on any of the questions asked (*Q*1 $t = .922$, *Q*2 $t = .588$ *Q*3 $t = .1.228$ *Q*4 $t = -.571$, all *p*'s > .1). Therefore, the survey did not reveal any liking or motivational differences based on fluency. However, follow-up analyses showed that the measure was indeed sensitive to differences between classes (e.g. between chemistry and history) in liking ($F(5153) = 4.965$, $p < .001$) and frequency of confusion ($F(5152) = 7.879$, $p < .001$). Thus the lack of observed liking/motivational differences between fluency conditions is unlikely to be due to insensitive measures.

## 6. Discussion

This study demonstrated that student retention of material across a wide range of subjects (science and humanities classes) and difficulty levels (regular, Honors and Advanced Placement) can be significantly improved in naturalistic settings by presenting reading material in a format that is slightly harder to read. While disfluency appears to operate as a desirable difficulty, presumably engendering deeper processing strategies (c.f. Alter et al., 2007), the effect is driven by a surface feature that prima facie has nothing to do with semantic processing.

One alternative explanation to the notion that disfluency leads to deeper processing is that the hard-to-read fonts were more distinctive and that the effects were driven by distinctiveness. While we cannot conclusively rule out this possibility, it is worth noting that we sampled fonts that are within the normal variation used in textbooks and classroom environments. So, while the disfluent fonts were less typical than our fluent fonts, they were not extreme as to cause them to stand out as unusual. Moreover, over the course of a semester, any novelty of the hard-to-read fonts should wear off thus reducing the impact of distinctiveness. Of course, it is likely the effects are multiply determined, which makes pinning down the precise mechanism quite challenging. Regardless of the underlying cognitive process, the implications of the finding for educational practice are non-trivial.

However, while these results are an encouraging sign of the promise of cognitive educational interventions, there are several caveats to consider. It is important to ascertain the point at which material is no longer disfluent, but instead illegible, or otherwise unnecessarily difficult to the point that it hinders learning. It seems possible that the influence of disfluency on retention follows a U-shaped curve, and the exact parameters of this function remain to be determined. With this in mind, the most effective disfluency manipulations would likely be those that are within the bounds of the normal variation of fonts and materials that could reasonably appear in a classroom.

It is also worth noting that fluency manipulations can come in many forms, some of which may be less likely to lead to improvements in retention. For example, in an elegant set of studies, Rhodes and Castel (2008) demonstrated biases in metacognition by manipulating fluency via font size. Larger fonts led participants to believe that they would have better recall, but in reality memory did not differ as a function of font size. Why did not small fonts lead to better retention, as the disfluency account would predict? One reason might be that the small font was 18 point Arial (the large font was 48 point Arial). Even their small font was considerably larger than standard. While the 18 point font was relatively less fluent than the 48 point font, it still might not have been disfluent enough to bring about the deeper thinking that disfluency engenders (c.f. Alter et al., 2007). Further studies will be necessary to titrate optimal levels of disfluency.

Other forms of desirable difficulties have been shown to be moderated by factors such as the nature of the materials (McDaniel, Hines, & Guynn, 2000) and how the materials are tested (Thomas & McDaniel, 2007). The present studies examined naturalistic materials and testing strategies across a wide array of topics, but were hardly exhaustive of the space of educational materials. Moreover, less motivated or able students from less successful schools might be more inclined to give up on the material rather than persist and encode it more deeply (c.f. McNamara, Kintsch, Butler-Songer, & Kintsch, 1996). Thus, further research is advisable before widespread implementation of disfluency interventions.

That said, fluency interventions are extremely cost-effective, and font manipulations could be easily integrated into new printed and electronic educational materials at no additional cost to teachers, school systems, or distributors. Moreover, fluency interventions do not require curriculum reform or interfere with teachers' classroom management or teaching styles.

The potential for improving educational practices through cognitive interventions is immense. If a simple change of font can significantly increase student performance, one can only imagine the number of beneficial cognitive interventions waiting to be discovered. Fluency demonstrates how *small interventions* have the potential to make big improvements in the performance of our students and education system as a whole.

## Acknowledgements

Thanks to Dr. Beth Yauman, a Princeton Undergraduate research fellowship and the PSURE research fellowship for supporting the first and third authors respectively. Thanks to Robert Bjork, Rolf Reber, an anonymous reviewer, Michelle Ellefson, Alan Castel, Adam Alter, David Schkade, Ani Momjian, Anuj Shah, Jeff Zemla, and the Opplab for advice and support.

## References

Alter, A. L., & Oppenheimer, D. M. (2008). Effects of fluency on psychological distance and mental construal (or why New York is a large city, but New York is a civilized jungle). *Psychological Science, 19*, 161–167.

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

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology, 136*(4), 569–576.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Book, C., Byers, J., & Freeman, D. (1983). Student expectations and teacher education traditions with which we can and cannot live. *Journal of Teaching Education, 34*(9), 1–9.

Brophy, J. (1983). Research on self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*(5), 631–661.

Castel, A. D., McCabe, D. P., & Roediger, H. L. III, (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin and Review, 14*, 107–111.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech effect language comprehension. *Cognition, 105*, 658–668.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*(6), 671–684.

Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104*(3), 268–294.

Haynes, N., & Johnson, S. (1983). Self- and teacher expectancy effects on academic performance of college students enrolled in an academic reinforcement program. *American Educational Research Journal, 20*(4), 511–516.

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

McDaniel, M. A., Hines, R., & Guynn, M. (2000). When text difficulty benefits less-skilled readers. *Journal of Memory and Language, 46*(3), 544–561.

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

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences, 12*(6), 237–241.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science, 9*, 45–48.

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.

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, & M. Bucciarelli (Eds.), *Proceedings of the twenty-seventh annual conference of the cognitive science society*. Mahwah, NJ: Erlbaum.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *JEP: Human Learning and Memory, 4*, 592–604.

Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition, 26*(6), 791–799.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233.

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