

# On Writing Legibly: Processing Fluency Systematically Biases Evaluations of Handwritten Material

Social Psychological and  
Personality Science  
1(3) 230-237  
© The Author(s) 2010  
Reprints and permission:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1948550610368434  
<http://spps.sagepub.com>



Rainer Greifeneder<sup>1,2</sup>, Alexander Alt<sup>2</sup>, Konstantin Bottenberg<sup>2</sup>,  
Tim Seele<sup>2</sup>, Sarah Zelt<sup>2</sup>, and Dietrich Wagener<sup>2</sup>

## Abstract

Evaluations of handwritten essays or exams are often suspected of being biased, such as by mood states or individual predilections. Although most of these influences are unsystematic, at least one bias is problematic because it systematically affects evaluations of handwritten materials. Three experiments revealed that essays in legible as compared to less legible handwriting were evaluated more positively. This robust finding was related to a basic judgmental mechanism that builds on the fluency with which handwriting can be processed. The present research further revealed that this evaluative bias is not inevitable but can be controlled for. Given the importance of evaluations based on handwritten work samples for individual success throughout school, college, university, and work life, it is important for individuals to be aware of this bias.

## Keywords

handwriting, legibility, evaluation, fluency, education

Both those who grade exams and those whose exams are graded have long suspected that evaluations are not always objective but biased by a host of influences, including mood states and individual predilections. Most of these biases are unsystematic because they vary across time and graders and may thus not be particularly worrisome on an aggregate level. However, at least one bias is problematic because it systematically affects evaluations of handwritten material. Specifically, handwriting legibility biases evaluation, with essays in legible as compared to less legible handwriting being evaluated more positively (e.g., James, 1929).<sup>1</sup> This bias—hereafter referred to as legibility bias—can be highly consequential because individual success throughout school, college, university, and work life depends only in part on standardized multiple-choice or computer-based performance assessments and often on evaluations of handwritten materials such as essays, exams, or even handwritten résumés, such as in France. Understanding the source of this bias, and ways to control it, therefore appears critical. The present research seeks to fulfill these goals, thus moving the field from acknowledging the legibility bias to understanding its cause and cure.

Asked about the source of this bias, laypersons' explanation seems based on the assumption that handwriting is indicative of personality, as handwriting apparently comprises the wealth of characteristics needed to mirror trait differences (e.g., Ben-Shakhar, Bar-Hillel, Bilu, Ben-Abba, & Flug, 1986). Yet graphology ratings have virtually zero predictive validity

(Neter & Ben-Shakhar, 1989) and, even more importantly here, are not reliable across raters (e.g., Bayne & O'Neill, 1988). Because of this, graphology-like inferences cannot account for the observed *systematic* effect of legibility on evaluations.

Another explanation for why legible handwritten materials receive higher grades is the consideration of penmanship. James (1929) reported that essays in legible versus less legible handwriting were evaluated more positively. Presumably this was because evaluators awarded a premium for legible handwriting or penalized less legible handwriting, as good penmanship was a virtue at that time. Similar reasoning may explain later replications, in which elementary school teachers—who teach penmanship and spontaneously take its mastery into account—were recruited as participants (e.g., Briggs, 1970; Markham, 1976) or in which essay topics such as “Hopes and Aspirations for the Next Decade”—which lack objective content criteria and may therefore invite the consideration of penmanship—were used (e.g., Hughes, Keeling, & Tuck, 1983; Klein & Taub, 2005). Nowadays, however, it would

<sup>1</sup> Radboud University Nijmegen, Netherlands

<sup>2</sup> University of Mannheim, Germany

## Corresponding Author:

Rainer Greifeneder, Radboud University Nijmegen, Behavioral Science Institute, Postbus 9104, 6500 HE Nijmegen, Netherlands  
Email: [greifeneder@uni-mannheim.de](mailto:greifeneder@uni-mannheim.de)

appear that mastery of penmanship is less important, at least beyond elementary school and when objective content criteria are available. If evaluations are still biased by differences in legibility, penmanship is unlikely to be the culprit.

As a new explanation, we propose that the legibility bias results from a basic judgmental mechanism that takes the fluency of information processing into account. Specifically, we suggest that individuals spontaneously form inferences such as, "If it can be processed fluently, it is probably good." These inferences are then used as information in evaluation. This new explanation is much broader than prior theorizing because it suggests that the legibility bias is a pervasive phenomenon. What follows seeks to substantiate this fluency hypothesis.

Fluency refers to the felt ease or difficulty with which mental processes can be executed (e.g., Greifeneder & Bless, in press). A large body of evidence suggests that individuals spontaneously recruit feelings of fluency when forming judgments of various kinds (for overviews, see, e.g., Alter & Oppenheimer, 2009; Schwarz & Clore, 2007). For instance, statements are endorsed as more probably true when processing is fluent (e.g., Reber & Schwarz, 1999; for a review, see Dechêne, Stahl, Hansen, & Wänke, in press), liking is enhanced when stimuli are processed fluently (e.g., Reber, Winkielman, & Schwarz, 1998), and events are judged as more frequent and products as more positive when recollection is easy (e.g., Greifeneder & Bless, 2008; Wänke, Bohner, & Jurkowitsch, 1997). The diversity of these examples suggests that the influence of fluency on judgment is a pervasive phenomenon (also see Alter & Oppenheimer, 2009).

But why is the influence of fluency on judgments pervasive? Supposedly, this is because fluency is generally perceived to be "about" whatever is the focus of attention (Clore et al., 2001; Higgins, 1996), even if such attributions are not warranted. Once attributed, fluency is usually interpreted as a signal of positivity, and disfluency as a signal of negativity (Schwarz, Song, & Xu, 2008), because, over time, individuals have learned that positive states of the world are associated with processing fluency and negative states with disfluency (Unkelbach, 2006, 2007). Consistent with this evidence, we suggest that individuals spontaneously attribute fluency when processing handwritten material to the focus of attention, such as the handwritten material or its author. Fluency is then used as information to draw inferences about the positivity or negativity of these attribution targets. The present contribution seeks to substantiate this new explanation by demonstrating (a) that the legibility bias occurs even if penmanship is not a concern, (b) that it is related to processing fluency, and (c) that it can be controlled for by drawing participants' attention to its source.

## Experiment 1

### Method

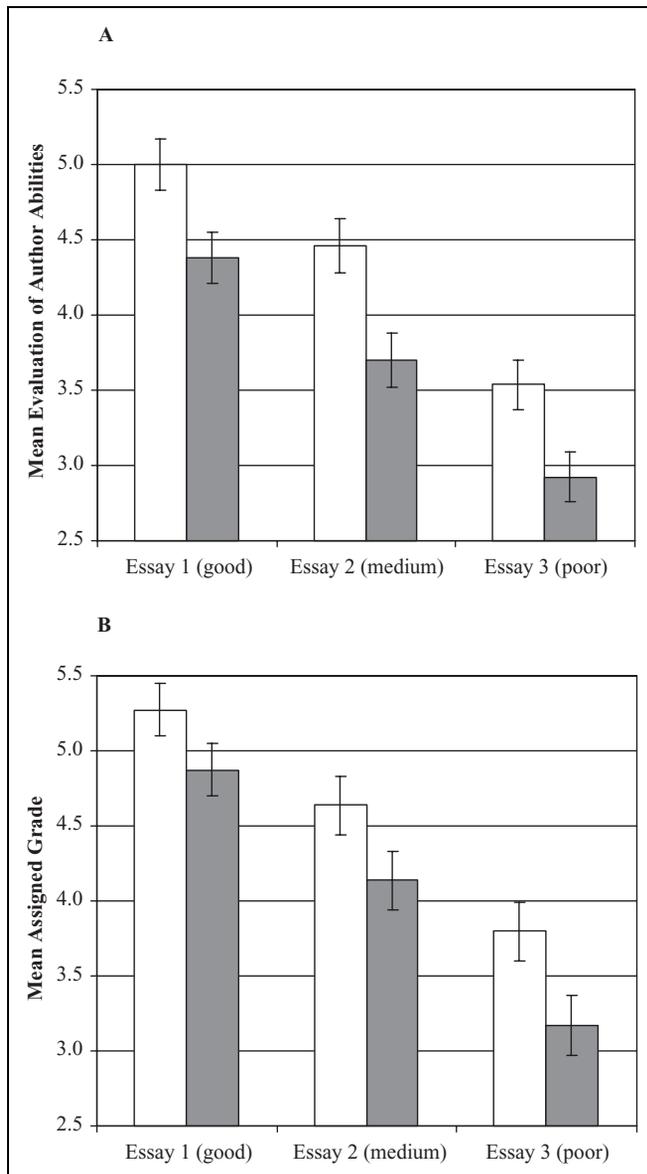
**Participants and design.** A total of 44 University of Mannheim students (22 female; age  $M = 23.14$  years,  $SD = 2.89$ ) received 2 EUR and a chocolate bar to evaluate a good, medium, and

**Table 1.** Design of Experiment 1

	Essay content quality		
	Good (Essay 1)	Medium (Essay 2)	Poor (Essay 3)
Legibility of essays in Group 1	Low	High	Low
Legibility of essays in Group 2	High	Low	High

poor essay. As we wanted all participants to grade at least one legible essay and one less legible essay, participants were randomly divided into two experimental groups. Some participants were presented the good, medium, and poor essays in, respectively, low, high, and low legibility. Others were presented the same essays in the same order but in high, low, and high legibility. This resulted in a 3 (content quality: good, medium, poor)  $\times$  2 (group: low-high-low vs. high-low-high legibility of essays) mixed-factorial design, with content quality as the within factor (see Table 1 for a concise overview of this design). As each of the three essays was presented in both highly and less legible handwriting, the hypothesized effect of legibility should be apparent when comparing the evaluations for *each* essay between groups. Importantly, because the assignment of legibility to essays was reversed between experimental groups (low-high-low vs. high-low-high), the direction of the hypothesized legibility effect should alternate from essay to essay. Accordingly, across the three essays, the hypothesized effect of legibility should be apparent in a significant interaction effect. Note that the present setup of evaluating several essays closely matches the situation graders usually encounter. Moreover, this setup reduces the likelihood that effects are due to handwriting features other than the systematically varied legibility.

**Essay construction.** Initially, a series of typed essays of similar length was constructed, with essays varying in the amount of correct information (content quality). Based on independent pretesting ( $N = 28$ , 12 female), three essays of good, medium, and poor content quality were selected (6-point standard German grade scale; all pairwise  $|ts| > 2.10$ ,  $ps < .04$ ). A new sample of students then copied these essays in their usual cursive handwriting, each on a separate blank sheet of paper. In second independent pretesting ( $N = 32$ , 18 female), participants evaluated handwriting on a scale from 1 (*easy to read*) to 6 (*difficult to read*). For each essay a highly and a less legible version was selected (all  $|ts| > 2.30$ ,  $ps < .05$ ), with the constraint that all words in all essays be readable. In a final independent pretesting ( $N = 27$ , 16 female), which followed the above described 3  $\times$  2 mixed-factorial design, participants were first asked to read the handwritten essays in their normal reading speed while processing latencies were recorded and then to revisit each essay and evaluate handwriting legibility. Results revealed that for each of the three essays the highly legible version not only was perceived as more legible (all  $|ts| > 2.80$ ,  $ps < .01$ ) but could also be read significantly faster (all  $|ts| > 2.08$ ,  $ps < .05$ ).<sup>2</sup>



**Figure 1.** Mean evaluations (with standard errors) of author abilities (1a) and assigned grades (1b) for each essay in Experiment 1. Note: Higher ratings indicate more positive evaluation. Means are not arranged by experimental group but rearranged by handwriting legibility. Evaluations of legible essays are displayed as white bars, evaluations of less legible essays as black bars.

**Procedure.** On arrival at the laboratory, participants were handed a questionnaire including all instructions and materials. To foster a high level of accuracy motivation and thereby decrease evaluation bias, participants were told that the experiment focused on interrater reliability. Participants read a short paragraph (74 words) about a physics topic, “The Emergence of Lightning,” and were asked to evaluate the subsequent essays based on this standard. The essays were supposedly written by students as part of a teaching assignment. Participants read and evaluated the essays one by one, starting with an example before working on the three target essays.

**Evaluation.** After each essay, participants were asked to evaluate the presumed author regarding general academic competence, knowledge of other school subjects, diligence, time spent studying, verbal expressiveness, and abilities in other domains.<sup>3</sup> All six evaluations were assessed on 6-point Likert-type scales (1 = *high* to 6 = *low*). Subsequently, participants assigned a grade to the respective essay on a scale from 1 (*excellent*) to 6 (*insufficient*; standard German grade scale).

## Results and Discussion

Both evaluations of author abilities and assigned grades were individually rescaled such that higher scores indicate more positive evaluations. For each essay, the six items targeting author abilities were highly interrelated and averaged (all Cronbach’s  $\alpha > .87$ ). The indices for author abilities and assigned grades were separately subjected to  $3 \times 2$  mixed-factorial ANOVAs.

For *author abilities*, a main effect of content quality indicates that presumed authors were evaluated more positively the higher the quality of the essays, reflecting that participants took content quality into account,  $F(2, 84) = 45.34, p < .01, \eta^2 = .52$ . More importantly, across essays, highly legible as compared to less legible handwriting led to more positive evaluations, as apparent in the hypothesized interaction effect,  $F(2, 84) = 13.34, p < .01, \eta^2 = .24$  (main effect group,  $F < 1$ ). Planned comparisons indicated that this legibility effect is strong for each of the three essays,  $t(42) = 2.54, p < .02$ , Cohen’s  $d = 0.77$ ;  $t(42) = 2.99, p < .01$ , Cohen’s  $d = 0.90$ ;  $t(42) = 2.62, p < .02$ , Cohen’s  $d = 0.79$ , thus ruling out alternative explanations that assume differences for certain levels of content quality only. The cell means are depicted in Figure 1a; for ease of presentation, means are not arranged by experimental group but are rearranged by handwriting legibility.

A parallel pattern of evidence was observed for *assigned grades*, with essays evaluated more positively the better they were content-wise,  $F(2, 84) = 43.93, p < .01, \eta^2 = .51$ , but also the more legible the handwriting was, as reflected in a significant interaction effect,  $F(2, 84) = 6.10, p < .01, \eta^2 = .13$  (main effect group,  $F < 1$ ). Again, it is noteworthy that a considerable percentage of the variance in grade ratings was because of systematic differences in legibility. Note also that, on average, highly legible essays were evaluated 0.5 grade points more positively than less legible essays, which should be considered serious given a 6-point grade scale. Individual cell means are depicted in Figure 1b.

In sum, Experiment 1 revealed that handwriting legibility systematically biases evaluations of both author abilities and content quality. This occurred even though penmanship was unlikely to be a concern because the sample consisted of students (instead of, e.g., elementary school teachers), and these students evaluated essays based on explicit content criteria as provided in the standard paragraph. It was further shown that the effect occurred for good, medium, and poor essays and was significant both between and within groups, thus attesting to the legibility bias’ general nature.

Experiment 1 relied on a student sample to reduce the potential impact of penmanship considerations. As a downside of this choice, the results may be perceived as less ecologically valid. To address this concern, a follow-up study was conducted, which employed the same design, procedure, and materials as in Experiment 1 but recruited a convenience sample of 47 German secondary school teachers, who regularly evaluate handwritten exams. Replicating the findings observed in Experiment 1, highly legible as compared to less legible handwriting resulted in significantly more positive evaluations and assigned grades. This finding further attests to the legibility bias' general significance.

## Experiment 2

Experiment 1 revealed that essays in legible as compared to less legible handwriting were evaluated more positively. We suggest that this is because legible essays can be processed more fluently. To empirically substantiate this hypothesis, we also measured perceived legibility as a proxy for fluency. Furthermore, we wanted to rule out the alternative explanation that legible essays are evaluated more positively because they are perceived as more beautiful, reflecting a what is beautiful is good heuristic (Dion, Berscheid, & Walster, 1972). Indeed, one could suppose that legible essays are evaluated more positively because legibility and beauty often go together. Although a posttest suggests that this was not the case in Experiment 1, it appeared desirable to address this concern experimentally. Accordingly, new handwritings were selected for Experiment 2 such that legibility and beauty varied orthogonally. We expected that only legibility but not beauty would influence evaluations.

## Method

**Participants and design.** A total of 84 University of Mannheim students (68 female, 4 unknown; age  $M = 21.54$  years,  $SD = 4.21$ ) received 2 EUR and a chocolate bar to evaluate three essays of good, medium, and poor quality. Replicating the design of Experiment 1, some participants were presented the good, medium, and poor essay in, respectively, low, high, and low legibility. Other participants were presented the same essays in the same order but in high, low, and high legibility. In addition, handwriting beauty was varied between conditions: Some participants were presented with beautiful handwriting only, whereas other participants evaluated only materials in less beautiful handwriting.<sup>4</sup> This resulted in a 3 (content quality: good, medium, poor)  $\times$  2 (group: low-high-low vs. high-low-high legibility of essays)  $\times$  2 (handwriting beauty: high vs. low) mixed-factorial design, with content quality as within factor. Participants were randomly assigned to the four conditions. Materials and procedures were similar to Experiment 1, except for the following two changes.

**Essay construction.** A new sample of students was asked to copy the physics essays in their usual cursive handwriting, each

on a separate blank sheet of paper. In independent pretesting, participants ( $N = 9$ , 6 female) evaluated each handwriting with respect to legibility (1 = *easy to read* to 6 = *difficult to read*) and beauty (1 = *beautiful* to 6 = *unsightly*). Handwritings were then picked to represent the four following styles: high legibility-beautiful, high legibility-less beautiful, low legibility-beautiful, and low legibility-less beautiful. Highly legible versus less legible handwritings were rated as more legible but not as more beautiful,  $F(1, 8) = 113.98$ ,  $p < .01$ ,  $\eta^2 = .93$  (all other  $p > .17$ ), and beautiful versus less beautiful handwritings were rated as more beautiful but not as more legible,  $F(1, 8) = 65.64$ ,  $p < .01$ ,  $\eta^2 = .89$  (all other  $p > .12$ ).

**Assessment of perceived legibility.** After reading and evaluating all essays, participants in Experiment 2 were asked to look a second time at every essay and to evaluate the respective handwriting on a scale from 1 (*easy to read*) to 6 (*difficult to read*).

## Results and Discussion

Evaluations of author abilities were individually rescaled so that higher values indicate more positive evaluations, were averaged per essay (all Cronbach's  $\alpha > .83$ ), and were subjected to a  $3 \times 2 \times 2$  mixed-factorial ANOVA. Replicating Experiment 1, evaluations of author abilities were strongly contingent on the essays' content quality, with better essays being evaluated more positively,  $F(2, 160) = 122.63$ ,  $p < .01$ ,  $\eta^2 = .61$ . Moreover, essays in highly legible as compared to less legible handwriting were evaluated more positively, as reflected in a significant Legibility  $\times$  Content two-way interaction,  $F(2, 160) = 5.39$ ,  $p < .01$ ,  $\eta^2 = .06$ . Importantly, handwriting beauty did not significantly influence evaluations of author abilities,  $F(1, 80) = 1.22$ ,  $p > .27$  (all other  $F_s < 1$ ). If anything, beautiful versus less beautiful handwritings were associated with *lower* evaluations of author abilities ( $M = 3.93$ ,  $SD = 0.60$ ;  $M = 4.08$ ,  $SD = 0.59$ ), thus refuting the alternative hypothesis that essays in legible handwriting are evaluated more positively because what is beautiful is often also good. The 12 cell means are displayed in Table 2 (for ease of presentation, means are not arranged by experimental group but are rearranged by legibility).

To more directly test the proposed fluency hypothesis, mediation analyses were performed following Baron and Kenny (1986). Specifically, separately for each of the three essays but collapsing across beauty conditions, we analyzed whether the effect of the legibility manipulation (independent variable) on evaluations of author abilities (dependent variable) is mediated by perceived legibility, which supposedly constitutes a good proxy for fluency. Strong mediation was observed for each of the three essays (Sobel's  $z = 2.23$ ,  $p < .03$ ;  $z = 3.28$ ,  $p < .01$ ;  $z = 2.53$ ,  $p < .02$ ), suggesting that handwriting legibility biases evaluations because fluency associated with legibility is used as information when forming judgments.

## Experiment 3

Experiments 1 and 2 demonstrated that essays in legible handwriting are evaluated more positively, and this was linked to

**Table 2.** Mean Author Evaluations (With Standard Deviations) in Experiment 2 as a Function of Content Quality, Legibility of Essays, and Handwriting Beauty

	Essay content quality					
	Good (Essay 1)		Medium (Essay 2)		Poor (Essay 3)	
	M	SD	M	SD	M	SD
Beautiful handwriting						
Highly legible essays	5.08	0.81	3.93	0.92	3.17	0.79
Less legible essays	4.76	0.62	3.44	0.79	3.22	0.85
Less beautiful handwriting						
Highly legible essays	5.21	0.52	4.11	1.06	3.39	0.98
Less legible essays	4.78	0.80	3.90	0.97	3.08	0.87

Note: Author evaluations were assessed on 9-point Likert-type items. Higher values indicate more positive evaluation. Means are not arranged by experimental group but rearranged by handwriting legibility.

spontaneous inferences based on processing fluency. Experiment 3 sought to test whether this basic judgmental mechanism can be controlled for if participants know about the biasing influence of legibility. Such a possibility would provide an important means to counter the legibility bias and would further attest to the proposed fluency hypothesis, as detailed below.

## Method

A total of 108 University of Mannheim students (57 female; age  $M = 22.26$  years,  $SD = 3.40$ ) participated in return for 1.50 EUR and a chocolate bar. Of these, 2 participants had insufficient knowledge of the German language and 6 had already participated in one of the previous studies. These 8 participants were excluded from further analyses.

Participants were randomly assigned to a 3 (content quality: good, medium, poor)  $\times$  2 (group: low–high–low vs. high–low–high legibility of essays)  $\times$  2 (control vs. information) mixed-factorial design, with content quality as within factor. Half of the participants received the same instructions as in the previous experiments (control); the other half (information) also read, “Please note: Prior research revealed that the ease or difficulty with which handwritten essays can be read strongly impacts their evaluation. Please try not to be influenced by how easy or difficult it is to read the following essays.” The second sentence was subsequently repeated above each essay. All other materials and procedures replicated those of Experiment 1, though the number of dependent variables per essay was reduced from six to four for reasons of test efficiency. Because individuals generally cease to rely on fluency in judgment when fluency is said to be undiagnostic (e.g., Schwarz et al., 1991), we hypothesized that the influence of legibility on evaluations would be reduced or eliminated in the information condition. Together, these predictions translate to an expected three-way interaction.

## Results and Discussion

Evaluations of author abilities and grades were individually rescaled so that higher values indicate more positive

evaluations, were averaged per essay (all Cronbach’s  $\alpha > .82$ ), and were subjected to a  $3 \times 2 \times 2$  mixed-factorial ANOVA. This analysis revealed a significant main effect of content quality,  $F(2, 192) = 109.51, p < .01, \eta^2 = .53$ , a significant Content Quality  $\times$  Group two-way interaction,  $F(2, 192) = 4.53, p < .02, \eta^2 = .05$ , a significant Content Quality  $\times$  Information Condition two-way interaction,  $F(2, 192) = 3.18, p < .05, \eta^2 = .03$ , and a significant main effect of experimental group,  $F(1, 96) = 5.90, p < .02, \eta^2 = .06$ . Importantly, all of these effects were qualified by the expected significant three-way interaction,  $F(2, 192) = 5.96, p < .01, \eta^2 = .06$ . The 12 cell means are displayed in Table 3.

To further investigate this pattern of results, the control condition and the information condition were analyzed separately. Replicating prior findings, participants in the *control* condition (no additional information) assigned more positive evaluations the higher the essays’ content quality,  $F(2, 100) = 39.65, p < .01, \eta^2 = .44$ , and the higher the essays’ legibility, as reflected in a significant two-way interaction,  $F(2, 100) = 10.65, p < .01, \eta^2 = .18$  (main effect group,  $F < 1$ ). Planned comparisons indicated that for every essay, the highly legible version led to more positive evaluations than the less legible version,  $t(50) = 1.91, p < .07$ , Cohen’s  $d = 0.54$ ;  $t(50) = 3.28, p < .01$ , Cohen’s  $d = 0.91$ ;  $t(50) = 2.58, p < .02$ , Cohen’s  $d = 0.72$ , for the three essays, respectively.

Participants in the *information* condition also assigned more positive evaluations the better the essays were content-wise, as reflected in a significant content main effect  $F(2, 92) = 72.34, p < .01, \eta^2 = .61$ . Importantly, however, legible as compared to less legible essays did not consistently result in more positive evaluations, as indicated by a nonsignificant interaction term,  $F < 1$ . Thus, legibility did not systematically bias evaluations in the information condition, presumably because participants, once informed about the deleterious impact of handwriting legibility, were apt to correct for this bias. Unexpectedly, essays in the high–low–high group led to more positive evaluations than essays in the low–high–low group ( $M = 4.28, SD = 0.53$ ;  $M = 3.91, SD = 0.53$ ), as reflected in a significant main effect of experimental group,  $F(1, 46) = 5.58, p < .03, \eta^2 = .11$ . Further planned comparisons proved reliable only for the second essay,

**Table 3.** Mean Evaluations (With Standard Deviations) in Experiment 3 as a Function of Content Quality, Legibility of Essays, and Information Condition

	Essay content quality					
	Good (Essay 1)		Medium (Essay 2)		Poor (Essay 3)	
	M	SD	M	SD	M	SD
Control (no additional information)						
Highly legible essays	4.76	0.63	4.62	0.44	3.70	0.81
Less legible essays	4.41	0.69	4.01	0.86	3.11	0.84
Additional information						
Highly legible essays	4.98	0.75	4.08	0.84	3.35	0.80
Less legible essays	4.69	0.63	4.50	0.72	2.98	0.90

Note: Evaluations were assessed on 9-point Likert-type items. Higher values indicate more positive evaluation. Means are not arranged by experimental group but rearranged by handwriting legibility.

$t(46) = 1.87, p < .07$  (all other  $ps > .13$ ), and for this second essay, lower legibility led to more (and not less) positive evaluations. Potentially, this is because participants not only corrected but overcorrected for the supposed influence of fluency, resulting in the opposite of the generally observed legibility bias (for evidence on overcorrection, see, e.g., Strack, Schwarz, Bless, Kübler, & Wänke, 1993). Together, these results spark hope that the legibility bias may be successfully countered by instructing individuals to do so.

Although the primary objective of Experiment 3 was to identify a cure, the observed results also attest to the validity of the fluency hypothesis. The logic is as follows: If directing participants' attention to fluency reduces the effect, one may conclude that fluency was responsible for the observed legibility bias in the first place—for why else would the effect be reduced when participants are aware of the presently undue influence of fluency? Experiment 3 thus further attests to the suggested fluency hypothesis by way of moderation evidence (for details on this reasoning, see Schwarz et al., 1991; Spencer, Zanna, & Fong, 2005).

## General Discussion

Three experiments revealed that handwriting legibility systematically biases evaluations of author abilities and grades assigned to the respective essays. This effect was observed in contexts where penmanship is unlikely to be considered an important part of performance, for different levels of content quality (good, medium, and poor), and for different populations (students and secondary school teachers), thus ruling out a series of potential alternative explanations. Furthermore, the effect was not related to differences in handwriting beauty. Rather, essays in more legible handwriting were evaluated more positively because of the fluency associated with their processing. This conclusion dovetails with and extends findings in the domain of social and cognitive psychology, holding that the fluency associated with information processing influences judgments of various kinds and constitutes a frequent source of information in conditions of daily

life (e.g., Greifeneder, Bless, & Pham, in press; Schwarz et al., 2008).

The present research set out to investigate the process underlying the legibility bias. Although the observed evidence sustains the suggested fluency hypothesis, it should be acknowledged that the mediation in Experiment 2 was based on perceived rather than measured fluency and thus on a subjective measure. Although an objective measure would have been desirable, this appears secondary in light of the fact that the fluency hypothesis was supported with all three primary methodological approaches known to the field: (a) by refuting a series of alternative hypotheses, as in Experiments 1 and 2, (b) by testing mediation in Experiment 2, and (c) by testing moderation in Experiment 3.

It should also be noted that alternating highly legible and less legible handwriting may have helped fluency experiences to be salient (e.g., Hansen & Wänke, 2008). Yet this setup is likely to be of high ecological validity because essays or exams are rarely ordered by legibility, so that varying legibility levels is the very situation that graders usually encounter.

By relating the effect of legibility on evaluations to a fluency-based judgmental mechanism, the present findings suggest that the legibility bias is a pervasive phenomenon. This conclusion extends the scope of prior research on handwriting and points to a potentially harmful source of error because evaluations of handwritten material may be consequential at all stages of life. From this perspective, the final piece of evidence furnished by the present research is of particular relevance: Participants did not show the legibility bias when alerted to its existence, suggesting that the bias is not inevitable but can be controlled for.

This final piece of evidence also resolves a seeming discrepancy between the present findings and the notion of "wisdom in feelings" (e.g., Schwarz, 2002), which suggests that reliance on feelings is a generally sensible judgment mechanism—provided correct attribution (Schwarz, 2004). When appropriate attribution targets are unknown or unnoticed, however, feelings may lead astray because attribution is then primarily guided by temporal contiguity (e.g., Clore

et al., 2001). Presumably this is what causes the legibility bias, as fluency from legibility appears to be influential simply because it is perceived *when* forming evaluations. The present contribution thus sheds light on the Achilles's heel of reliance on fluency and emphasizes that for fluency feelings to be "wise," appropriate conceptions of causation are important. Such conceptions are available in certain domains and situations, for instance, with respect to undue influences of media coverage on perceived name frequency (Oppenheimer, 2004), but appear to be missing for legibility. For the sake of fair performance assessments based on handwritten material, it therefore appears critical that individuals know about legibility's potential for bias.

### Authors' Note

Authors 2 to 5 have contributed equally and are listed in alphabetical order.

### Declaration of Conflict of Interest

The author(s) declared that they had no conflicts of interests with respect to their authorship or the publication of this article.

### Financial Disclosure/Funding

The author(s) declared that they received the following financial support for their research and/or authorship of this article. This research was supported by a postdoctoral grant from the Deutsche Forschungsgemeinschaft and a postdoctoral grant from the Landesstiftung Baden-Württemberg, both awarded to the first author.

### Notes

1. Another systematic bias is the "beauty-is-talent" effect (Landy & Sigall, 1974), which holds that essays are evaluated more positively when authored by presumably attractive as compared to unattractive students.
2. Because of space limitations, only general results are reported for the pretests, Replication 1, and the mediation analysis in Experiment 2. Full results are available from the authors.
3. Asking for "diligence" may have been unfortunate because individuals potentially had a naïve theory that less readable handwriting directly reflects less diligence. There is reason to believe, however, that this alternative process did not produce the observed results. First, the same significance levels are obtained if author ability is computed without diligence. Second, a similar pattern of results is observed for grades. Third, evaluations in Experiment 3 are not influenced by differences in legibility once individuals are made aware of fluency, suggesting that fluency is the mediating link.
4. Varying legibility within participants but beauty between participants may be perceived as an unfair test of alternative hypotheses. Note, however, (a) that legibility is also varied between participants because each essay is presented in highly legible handwriting to some participants and less legible handwriting to other participants and (b) that the between-participants simple contrasts in Experiment 1 proved significant for every essay.

### References

- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219-235.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.
- Bayne, R., & O'Neill, F. (1988). Handwriting and personality: A test of some expert graphologists' judgements. *Guidance & Assessment Review, 4*(4), 1-3.
- Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y., Ben-Abba, E., & Flug, A. (1986). Can graphology predict occupational success? Two empirical studies and some methodological ruminations. *Journal of Applied Psychology, 71*(4), 645-653.
- Briggs, D. (1970). The influence of handwriting on assessment. *Educational Research, 13*(1), 50-55.
- Clore, G. L., Wyer, R. S., Dienes, B. P. A., Gasper, K., Gohm, C., & Isbell, L. (2001). Affective feelings as feedback: Some cognitive consequences. In L. L. Martin & G. L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 27-62). Mahwah, NJ: Lawrence Erlbaum.
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (in press). The truth about the truth: A meta-analytic review of the truth-effect. *Personality and Social Psychology Review*.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285-290.
- Greifeneder, R., & Bless, H. (2008). Depression and reliance on ease-of-retrieval experiences. *European Journal of Social Psychology, 38*(2), 213-230.
- Greifeneder, R., & Bless, H. (in press). The fate of activated information in impression formation: Fluency of concept activation moderates the emergence of assimilation versus contrast. *British Journal of Social Psychology*.
- Greifeneder, R., Bless, H., & Pham, M. T. (in press). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review*.
- Hansen, J., & Wänke, M. (2008). It's the difference that counts: Expectancy/experience discrepancy moderates the use of ease of retrieval in attitude judgments. *Social Cognition, 26*, 447-468.
- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133-168). New York, NY: Guilford.
- Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement, 20*(1), 65-70.
- James, H. W. (1929). The effect of handwriting upon grading. *English Journal, 16*(3), 180-185.
- Klein, J., & Taub, D. (2005). The effect of variations in handwriting and print on evaluation of student essays. *Assessing Writing, 10*, 134-148.
- Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29*(3), 299-304.

- Markham, L. R. (1976). Influences of handwriting quality on teacher evaluation of written work. *American Educational Research Journal*, 13(4), 277-283.
- Neter, E., & Ben-Shakhar, G. (1989). The predictive validity of graphological inferences: A meta-analytic approach. *Personality and Individual Differences*, 10(7), 737-745.
- Oppenheimer, D. M. (2004). Spontaneous discounting of availability in frequency judgment tasks. *Psychological Science*, 15(2), 100-105.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8(3), 338-342.
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, 9(1), 45-48.
- Schwarz, N. (2002). Situated cognition and the wisdom in feelings: Cognitive tuning. In L. F. Barrett & P. Salovey (Eds.), *The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 144-166). New York, NY: Guilford.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14(4), 332-348.
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, 61(2), 195-202.
- Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 385-407). New York, NY: Guilford.
- Schwarz, N., Song, H., & Xu, J. (2008). When thinking is difficult: Metacognitive experiences as information. In M. Wänke (Ed.), (pp. 201-223). New York, NY: Psychology Press.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89(6), 845-851.
- Strack, F., Schwarz, N., Bless, H., Kübler, A., & Wänke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology*, 23(1), 53-62.
- Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, 17(4), 339-345.
- Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 33, 219-230.
- Wänke, M., Bohner, G., & Jurkowitsch, A. (1997). There are many reasons to drive a BMW: Does imagined ease of argument generation influence attitudes? *Journal of Consumer Research*, 24(2), 170-177.

### Bios

**Rainer Greifeneder**, PhD, is a research associate at the Behavioral Science Institute of Radboud University and at the University of Mannheim.

**Alexander Alt** is a graduate student at the University of Mannheim.

**Konstantin Bottenberg** is a graduate student at the University of Mannheim.

**Tim Seele** is a graduate student at the University of Mannheim.

**Sarah Zelt** is a graduate student at the University of Mannheim.

**Dietrich Wagener**, PhD, is a lecturer at the University of Mannheim.