



The Impact of Orthographic Legibility on Writing Assessment: A Case Study of Secondary Students at a language school in Egypt

Andrew Wagdy Farag Ghaly

English Language Centre, Pharos University in Alexandria

Abstract- The assessment of second language (L2) writing is often influenced by rater-subjectivity factors that extend beyond linguistic competence. This study investigates the "legibility effect"—the impact of handwriting quality on evaluative outcomes—among 50 Senior 2 students at a language school in Alexandria, Egypt. Using a within-subjects design, student essays were graded once in their original handwritten form and once after being transcribed into a standardized typed format. Despite the use of an analytical rubric that excluded penmanship as a criterion, results from a paired-samples t-test revealed a significant increase in scores for the typed versions ($M = 38.16$, $SD = 5.12$) compared to the handwritten versions ($M = 33.42$, $SD = 7.84$); $t(49) = 5.38$, $p < .001$, $d = 0.72$. These findings suggest that poor legibility imposes a cognitive load on raters, triggering a "horns effect" that penalizes students, while neat handwriting creates a "halo effect" that may inflate scores. The study concludes that the transition to digital assessment in the Egyptian secondary system is a necessary step to ensure psychometric validity and educational equity.

Keywords: legibility effect, L2 writing assessment, halo effect, Senior 2, Egyptian education, rater bias.

I. Introduction

The evaluation of student writing remains one of the most intellectually demanding tasks in educational psychometrics, particularly within high-stakes environments like the Egyptian secondary system. In Alexandria, a city with a prestigious history of bilingual education, the intersection of paper-based testing and modern computer-mediated evaluation has exposed a persistent variable: the legibility effect. This phenomenon suggests that physical aesthetics can distort scores even when rubrics explicitly exclude penmanship.

The legibility effect is rooted in cognitive processing; poor handwriting creates a cognitive load that exhausts rater resources. Through processing fluency, easy-to-read text triggers a positive affective response (the halo effect), while messy text can trigger a negative bias (the horns effect). In the context of English as a Foreign Language (EFL) in Egypt, this effect is amplified by the Senior 2 stage's rigorous curriculum, where students must master complex structures and advanced vocabulary. As Egypt moves toward its "2030 Vision" for digital reform, understanding these biases is critical for ensuring the comparability of paper and tablet-based scores.

II. Review of literature

A substantial body of experimental work shows that more legible handwriting leads to higher grades for identical content. In a classic examination-script study, the same five English essays were recopied in five contrasting handwriting styles and marked by



practising teachers; analysis showed marks varied significantly not only by content but also by handwriting style, with differences large enough to shift a script from pass to fail (Briggs, 1980).

Klein and Taub found that handwriting legibility significantly affected teachers' evaluations of essays, with large effect sizes (Cohen's $d \approx 0.8-1.4$), such that legible versions received much higher marks than the same content in less legible scripts (Klein & Taub, 2005; Ruusuvirta et al., 2021).

Greifeneder and colleagues showed that legible essays were consistently evaluated more positively on both content quality and author ability, and received on average about 0.5 grade points higher on a 6-point scale compared with less legible versions of the same texts (Greifeneder et al., 2010; Greifeneder et al., 2012).

Teacher-expectancy research with spelling tests and essays found that illegible samples were graded more harshly: identical spelling lists with poor legibility were marked with more errors, and strong essays in illegible handwriting received lower scores than weak essays written legibly (Sprouse & Webb, 1994).

These findings are highly consistent with the current study's pattern: typed versions receiving significantly higher grades than the original handwritten scripts, even though the rubric nominally excluded handwriting. The shift from handwritten to typed format largely removes variation in letter formation, size, spacing, and alignment that underlie legibility judgements (Ziviani & Elkins, 1984; Graham et al., 1998; Graham et al., 2001; Sparaci et al., 2024).

A more recent field-based thesis with EFL university students reported that teachers overwhelmingly perceived poor handwriting as lowering students' test scores and described practical difficulties and negative attitudes toward illegible scripts (Houda & Houda, 2017).

However, Not all studies find a handwriting legibility effect; several show null or small effects, particularly with experienced examiners or when graders are alerted to the possibility of bias.

Ruusuvirta and others asked teachers to grade primary pupils' test answers that varied in content quality and handwriting legibility. Handwriting legibility did not significantly affect grades, although lower legibility reduced graders' confidence and was stereotypically associated with males (Ruusuvirta et al., 2021). The authors interpret this as spontaneous task-specific debiasing—graders noticing legibility as a potential bias and consciously compensating.

Earlier work with experienced examiners (e.g., Oxford University examiners) similarly reported no systematic bias from handwriting quality, or only small and inconsistent effects (Ruusuvirta et al., 2021).



A recent analysis of English essays found no clear significant correlation between an “illegible ratio” (proportion of difficult-to-read text) and rubric score; essay length and word count were much stronger predictors of grades (Dinh & Dinh, 2023).

Studies of medical undergraduates’ preliminary exam scores report that illegible handwriting did not significantly worsen test scores; low performance correlated more strongly with overall academic achievement rather than legibility alone (Chiwhane et al., 2020).

Importantly, experimental work has shown that explicit warnings about the legibility effect can reduce or eliminate it. Greifeneder and colleagues found that when raters were explicitly instructed about the threat of handwriting bias to content judgments, the legibility effect disappeared, indicating that awareness can trigger effective debiasing (Ruusuvirta et al., 2021; Greifeneder et al., 2010).

III. Method

Participants

The sample consisted of 50 students (N = 50) in the Senior 2 stage of a language school in Alexandria, Egypt. Participants were native Arabic speakers, aged 15–16, studying English as their primary foreign language.

Materials

A descriptive-argumentative essay prompt—“The Impact of Modern Technology on Stress and Time Management”—was used for the assessment. Essays were evaluated using a 50-point analytical rubric adapted from Egyptian Ministry of Education specifications. The rubric included five dimensions, each worth 10 points: Content and Ideas, Organization, Vocabulary, Grammar, and Mechanics. Notably, handwriting was not a criterion for grading.

Procedure

A within-subjects experimental design was employed across two phases. In Phase One, students wrote the essay on standard examination paper. These scripts were graded by the researcher. In Phase Two, the handwritten scripts were transcribed exactly into a standardized typed format (CBT mode), with no corrections to linguistics or mechanics. After a washout period to minimize rater memory, the typed versions were re-evaluated by the same rater using the same rubric.

IV. Results

Descriptive statistics indicated a noticeable shift in scores between the two assessment modes. The mean score for the typed condition was higher, and the scores showed less variability than in the paper-based condition.

Assessment Mode	Mean (M)	Standard Deviation (SD)
-----------------	----------	-------------------------



Paper-Based (Handwritten)	33.42	7.84
Computer-Based (Typed)	38.16	5.12

A paired-samples t-test was conducted to compare the scores of students in the handwritten and typed conditions. There was a significant difference in the scores for the handwritten ($M = 33.42$, $SD = 7.84$) and typed ($M = 38.16$, $SD = 5.12$) conditions; $t(49) = 5.38$, $p < .001$. The calculated effect size (Cohen's d) was 0.72, which represents a large effect. These results indicate that legibility had a significant impact on the grades awarded, with students receiving significantly higher scores when their work was presented in a legible, typed format.

V. Discussion

The findings provide empirical evidence of the legibility effect within the Alexandria language school context. The significant score increase in the typed condition suggests that raters unconsciously penalized poor penmanship in the paper phase, likely due to rater fatigue and the cognitive heuristics used to manage high grading volumes.

The Halo and Horns Effect

When work was typed, raters could more easily identify sophisticated vocabulary and correct grammar that were previously obscured. This confirms that "messiness" is often misread as a lack of mental organization. Conversely, students with neat handwriting may benefit from a halo effect on paper, where aesthetic appeal compensates for mediocre content. Once the "noise" of handwriting was removed, the evaluation became more centered on the linguistic and structural merits of the text.

Implications for Egyptian Education

The results have profound implications for the Thanaweya Amma examinations. If legibility accounts for such significant score variance, paper-based writing assessments may lack the necessary fairness for high-stakes decisions. The Ministry of Education's move toward tablet-based testing is an effective strategy to mitigate this bias. However, policymakers must be wary of "test mode effects," such as typing speed or system malfunctions, which could introduce new forms of inequity.

Recommendations and Limitations

To improve assessment reliability, schools should adopt blind grading and rater calibration sessions to identify rater-criteria interactions. Encouraging typed formative feedback can also help teachers focus on a student's actual linguistic competence. This study is limited by its sample size and the use of a single rater. Future research should involve multiple raters to explore inter-rater reliability and use Multi-Faceted Rasch Measurement (MFRM) to pinpoint exactly where the halo effect occurs across rubric criteria.



References

1. Briggs, D. (1980). A Study of the Influence of Handwriting upon Grades Using Examination Scripts. *Educational Review*, 32, 186-193. <https://doi.org/10.1080/0013191800320207>
2. Chihwane, A., Verma, A., & Chakole, S. (2020). ILLEGIBLE HANDWRITING AND PRELIMINARY EXAM TEST SCORES IN MEDICAL UNDERGRADUATES. **, 7, 1054-1060.
3. Dinh, R., & Dinh, A. (2023). The Discipline of Handwriting Quality and Essay Scoring. *Journal of Student Research*. <https://doi.org/10.47611/jsrhs.v12i4.5625>
4. Graham, S., Berninger, V., Weintraub, N., & Schafer, W. (1998). Development of Handwriting Speed and Legibility in Grades 1–9. *Journal of Educational Research*, 92, 42-52. <https://doi.org/10.1080/00220679809597574>
5. Graham, S., Weintraub, N., & Berninger, V. (2001). Which manuscript letters do primary grade children write legibly. *Journal of Educational Psychology*, 93, 488-497. <https://doi.org/10.1037/0022-0663.93.3.488>
6. Greifeneder, R., Alt, A., Bottenberg, K., Seele, T., Zelt, S., & Wagener, D. (2010). On Writing Legibly. *Social Psychological and Personality Science*, 1, 230 - 237. <https://doi.org/10.1177/1948550610368434>
7. Greifeneder, R., Zelt, S., Seele, T., Bottenberg, K., & Alt, A. (2012). Towards a better understanding of the legibility bias in performance assessments: the case of gender-based inferences.. *The British journal of educational psychology*, 82 Pt 3, 361-74. <https://doi.org/10.1111/j.2044-8279.2011.02029.x>
8. Houda, S., & Houda, N. (2017). The Impact of Poor Handwriting on EFL Students Test Scores Case of Teachers of English at Mohammed Kheider University.
9. Klein, J., & Taub, D. (2005). The effect of variations in handwriting and print on evaluation of student essays. *Assessing Writing*, 10, 134-148. <https://doi.org/10.1016/j.asw.2005.05.002>
10. Ruusuvirta, T., Sievänen, H., & Vehkamäki, M. (2021). Negative findings of the handwriting legibility effect: the explanatory role of spontaneous task-specific debiasing. *SN Social Sciences*, 1, 1-14. <https://doi.org/10.1007/s43545-021-00183-w>
11. Sparaci, L., Fantasia, V., Bonsignori, C., Provenzale, C., Formica, D., & Taffoni, F. (2024). Handwriting in primary school: comparing standardized tests and evaluating impact of grapho-motor parameters. *Reading and Writing*, 38, 1409 - 1434. <https://doi.org/10.1007/s11145-024-10562-3>
12. Sprouse, J., & Webb, J. (1994). The Pygmalion Effect and Its Influence on the Grading and Gender Assignment on Spelling and Essay Assessments..
13. Ziviani, J., & Elkins, J. (1984). An Evaluation of Handwriting Performance. *Educational Review*, 36, 249-261. <https://doi.org/10.1080/0013191840360304>