A comment on Granberg *et al.* (2025)

Erin Hengel*

October 2025

Granberg *et al.* (2025) make three further criticisms of Hengel (2022). First, they claim that non-classical measurement error is "hardly discussed [in Hengel (2022)] to any great extent" (Granberg *et al.*, 2025, p. 1). Second, they criticise Hengel (2022) for not showing a correlation between readability scores and other measures of quality. Third, they argue that readability scores are inappropriate proxies of "academic abstract quality".[1]

In my opinion, Granberg *et al.* (2025)'s first and second claims are unfair characterisations of Hengel (2022). I discuss non-classical measurement error in detail in Appendix D.2 (Hengel, Online Appendix, 2022, pp. 1–4). Although my own data reveal a weak correlation between readability scores and citations, other studies have found a stronger link using different proxies for academic quality; I review this literature in Appendix D.1 (Hengel, Online Appendix, 2022, p. 1):

> *Furthermore, numerous studies have validated readability scores against surrogate measures of reading comprehension. More readable high school and college-level correspondence courses have higher completion rates (Klare and Smart, 1973). More readable academic journals enjoy larger readerships (Richardson, 1977; Swanson, 1948); their most readable articles win more awards (Sawyer et al., 2008) and are downloaded more often (Guerini et al., 2012). More readable abstracts are also (generally) cited more frequently (see Dowling et al. (2018) and McCannon (2019) and Figure D.1). In a blog post, Lukas Püttmann compares abstract readability to page views of VoxEU.org columns: more readable columns are viewed three percent more often (Püttmann, 2017). Evidence from other studies linking readability and citations is, however, weaker (Berninger et al., 2017; Laband and Taylor, 1992; Lei and Yan, 2016). My own data suggest a positive relationship in papers published after 1990—and particularly those published post-2000—but no relationship before that (Figure D.1).*

Furthermore, Hengel (2022)'s Figure D.1 summarises correlations between readability scores and alternative measures of reading comprehension published in other studies; many of these studies find a positive relationship using advanced reading material for their textual input (Hengel, Online Appendix, 2022, p. 2).

In my opinion, the validity of Granberg *et al.* (2025)'s third claim in relation to Hengel (2022) depends on their definition of "academic abstract quality". If it refers to the overarching quality of an academic article, then I agree—readability scores are poor proxies of this concept. As emphasised in Hengel (2024), Hengel (2022) only evaluates the relationship between author gender and abstract readability, conditional on paper "quality", as proxied for by citations; it does not rely on—nor does it claim to rely on—an assumption that readability predicts scientific quality.[2] Indeed, my other work primarily uses citations—and has *never* used readability scores!—as an overarching proxy of academic quality (see, *e.g.*, Hengel and Moon, 2023; Alexander *et al.*, 2023; Doleac *et al.*, 2025).

* Brunel University of London, erin.hengel@gmail.com.

[1] Granberg *et al.* (2025) also ask that I notify `textstat`'s authors about the errors identified in Hengel, (2024), which I have done.

[2] In fact, I went to great lengths to clearly differentiate "readability" from other components of academic quality—see, for example Section 4.1 in Hengel (2022, pp. 2970–2973).

Alternatively, Granberg *et al.* (2025) may instead mean *writing quality* when they use the phrase "academic abstract quality". In Hengel (2022), I use the terms "readability", "reading comprehension" and "writing quality" interchangeably. This is confusing, especially as the latter concept maps imperfectly onto the former two—*e.g.*, text that is easy to understand may not be well written and vice versa. Moreover, if these errors systematically correlate with author gender, then estimates in Hengel (2022) are not informative about gender differences in writing quality without making additional (and probably untestable) assumptions about the nature and direction of this measurement error.[3] While I briefly address this issue (see Hengel, Online Appendix, 2022, p. 2), a more thoughtful discussion was probably warranted. Granberg *et al.* (2025)'s criticism on this point is well placed; I hope their paper reminds readers to interpret Hengel (2022)'s results carefully.

Granberg *et al.* (2025) conclude by showing gender differences in "abstract quality", where they proxy for the latter by training a BERT model to predict citation outcomes based on abstract text. I have re-run their code and largely reproduced their results. I am less familiar with BERT models than I am with readability scores, however, so I leave the assessment of these results to more informed experts.[4]

References

Alexander, D., Gorelkina, O., Hengel, E., & Tol, R. (2023). *Gender and the time cost of peer review.* Mimeo. Retrieved from https://erinhengel.github.io/time/time.pdf

Doleac, J., Hengel, E., & Mohnen, M. (2025). *The golden leap: gender differences in the Matthew effect.* Mimeo. Retrieved from https://erinhengel.github.io/matthew-effect/matthew-effect.pdf

Granberg, M., Jansson, J., & Yang, Y. (2025). *A short comment on "A response to Granberg et al. (2024)".* I4R Discussion Paper.

Hengel, E. (2022). *Online Appendix to "Publishing while female: Are women held to higher standards? Evidence from peer review".* Retrieved from https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/ej/132/648/10.1093_ej_ueac032/1/ueac032_online_appendix.pdf?Expires=1762265227&Signature=id4D4SI2Hr2~YpYwgpb1aLtD58WL6VBNNlYt66mOG7wr5ImLr7hBFRqPAoVYZQBb-UbyrKqxZP-6ajpeJz6LfPBGcIlpkpxJz2

Hengel, E. (2022). Publishing while female: Are women held ot higher standards? Evidence from peer review. *The Economic Journal, 132*(648), 2951–2991.

Hengel, E. (2024). *A response to Granberg et al. (2024).* I4R Discussion Paper.

Hengel, E., & Moon, E. (2023). *Gender and equality at top economics journals.* Mimeo. Retrieved from https://erinhengel.github.io/Gender-Quality/quality.pdf

[3] For example, suppose readability scores do not capture academic writing quality at all, but instead only pick up the simplicity of an idea. Then a gender gap in readability would emerge if women choose to research and write about ideas that are easier to formulate.

[4] Nevertheless, it is hardly surprising that Granberg *et al.* (2025) find a gender difference favouring women in their predicted citation outcome. As shown in Hengel and Moon (2023), gender differences in citations favour women, conditional on publication in a top-five economics journal—*i.e.*, conditional on analysing the exact same database as Hengel (2022). A BERT model trained to predict citation outcomes using abstract text from Hengel (2022) will therefore associate words in female-authored papers with higher citations, because these are precisely the papers in its training data with more citations (on average)!