Darby Tanner

Dr. Karnes

RHET 20073

16 October 2025

Methods Analysis Essay

The phrase “beauty is in the eye of the beholder” implies that quality is subjective. Many also claim that this sentiment applies to writing. Texts are deemed “good” according to the opinions of observers. But subjectivity is not scientific, which can stump students who are learning to write. Their texts might have perfect spelling and bulletproof grammar but still aren’t “strong.” What definable traits make a text successful, and how can students best learn these skills? In both of the studies I have selected from *Written Communication* for this analysis, researchers (1) identify features that define effective writing, (2) use objective measurements to assess text quality, and (3) make recommendations on how to improve writing instruction. The articles titles are as follows:

1. “Popularization Writing Skills Development: A Longitudinal Case Study of the Writing Process and Writing Outcomes in Nine Undergraduate Interdisciplinary Students” by Florentine Marnel Sterk, Merel van Goch, Michael Burke, and Iris van der Tuin.
2. “A Direct Functional Measure of Text Quality: Did the Reader Understand?” by Joachim Grabowski and Moti Mathiebe.

Both studies attempt to measure text quality and make judgments on how students should be taught to improve their writing skills, but how do their methods compare?

“Popularization Writing”

Sterk et al. argue that university students are ill-equipped to communicate research to general audiences even with years of training in academic writing. This ability to “communicate academic findings to a broad and nonexpert audience” is defined as *popularization discourse* (967). It’s vital that students learn popularization discourse during higher education because “popularization skills are useful, required even, in both academic and professional careers” (968). To support these claims, the authors performed a longitudinal case study following the development of popularization writing skills in nine Dutch undergraduate students over three years.

The authors structured their study around this **research question**: What do popularization writing skills look like for interdisciplinary students at the start and end of a 3-year undergraduate university program? (973). A longitudinal study is a great method for answering this question because it can track changes in writing processes and products over a long period of time. The authors can only prove their claim that popularization writing skills are lacking in college programs by observing how students learn to write over the course of their programs. Researchers developed qualitative criteria for text quality; collected evidence, including text drafts, keystroke logs, personal reflections, self-assessments, and interviews; and analyzed this evidence and assigned scores to each student’s observed competency in popularization discourse.

Participants were Dutch undergraduate students enrolled in three-year interdisciplinary programs of at least 24 courses (975). Because this study observes changes in the students’ writing *process* and final *product,* researchers performed two assessments — one in 2018 and the other in 2021. Both year’s assessments followed the same structure: students read a study on the “effects of night-time social media use on the mental health of teens” then spent one hour

crafting a 400-word newspaper article explaining the study for a general Dutch audience. A successful article was both interesting and understandable for the audience (975). Students were not trained in popularization discourse before taking the assessments in order to accurately determine if their education was sufficient.

While the assessments both followed similar structures, there were some key differences. In 2018, 140 Year 1 undergraduate students took the assessment and results were collected in person. Before starting the writing task, students took a 5-question Likert scale assessment about their preparation level and interest in the topic. They also received a briefing of the assigned article's research goals and the writing task from their seminar professors. This year, only the writing *product* was assessed, not the *process.* I found it difficult to understand exactly *how* the researchers measured text quality during 2018 because the details of the first study were not included. Researchers evaluated the writing *products* using Sterk and Van Goch's framework of 34 strategies for popularizing research across disciplines, which are grouped into five themes: Subject Matter, Tailoring Information to the Reader, Credibility, Stance, and Engagement (973-4). Exactly how they coded and measured the results from this year remains unclear.

In 2021, only nine of the original 140 students participated, and the assessment took place online. It's implied that COVID-19 impacted participation, and I believe that low participation drastically limited the robustness of the study's results. Ten computer programs were used during this assessment, most importantly, Inputlog, ActivePresenter, and Qualtrics for data collection. In plain language, the authors observed how students "planned, wrote, and revised their texts" using Inputlog and ActivePresenter, and coded the results based on existing writing models. These models include measurements such as word choice, sentence structure, cohesion, narrative

structure, etc. By assigning codes to certain measurements, the authors were able to judge each text according to quality standards and assign each text a grade for simple analysis.

First, Year 3 students used Qualtrics to complete a questionnaire about their education and experiences with both popularization discourse and academic discourse (976). Then, they received a debrief on Microsoft teams about the assignment, their memories of the first assessment in 2018, and a short interview about their experiences. Last, they completed the same writing assessment from 2018 "to enable the best circumstances for comparing results longitudinally" (975). The program Inputlog produced an XML file for each participant, which recorded their keystrokes and pauses. Inputlog failed to record for two participants, so researchers used the recorded video from ActivePresenter to manually review their writing processes.

This year, both the writing *process* and *product* were assessed. The authors' explanation Because writing involves cognitive effort, they examine their case study students' writing *processes* using Hayes's 2012 cognitive writing model (978). This model uses three main levels: resource, process, and control (971). They coded these levels as transcription, revision, source text, searching, formatting, planning, and pauses (977).

Authors compared writing samples from both years to see how students' developed their popularization discourse skills, organizing results into a rubric with assigned grades. The grades implied there was little improvement in skill over three years, with four students even regressing in skill. The authors were unable to determine if certain writing processes predicted writing quality because all participants used similar processes with differing results (986). The students engaged in a range of writing processes over the three years, but none of them showed much improvement in terms of learning popularization discourse from their programs (991). Therefore,

Sterk et al. gleaned that students must be explicitly taught popularization discourse as a genre in order for their' skills to improve enough to be effective in their future careers (992).

It's clear that I struggle to understand this study's exact methodology, and I would restructure the study in a few ways. At a high level, I understand their reasoning behind collecting multiple kinds of evidence (keystrokes, texts, and interviews), but the excessive amount of information made their methods convoluted and muddied their conclusions. Separate studies should have been enacted for analyzing the writing *process* and writing *product*. I found myself confusing the various writing models and forgetting the point of the research. In fact, all the time they spent collecting evidence for the writing *process* ended up inconclusive anyway.

I believe that their methodology to answer their research question did not need to involve research on the writing process. If we revisit the research question ("What do popularization writing skills look like for interdisciplinary students at the start and end of a 3-year undergraduate university program?"), the purpose is observing if students learned popularization skills in their degree programs, *not* analyzing how students' individual writing processes impact text quality. A follow-up study on process could add further context to their conclusions without weakening the current study.

"Direct Function"

Joachim Grabowski and Moti Mathiebe argue that text quality can be measured by asking one question: did the reader understand? It's difficult to measure text quality because different research fields (i.e., psychology, education, etc.) use different measurement criteria based on various theoretical frameworks, which may or may not yield "valid" results (204). In this article, the authors used a reference study that assessed the quality of individual texts from German

students at the beginning and end of secondary school (Grades 5 and 9). In their follow-up study, 277 university students attempt to recreate the accident on a computer program according to the descriptions produced by the secondary school students' reports. If the university students successfully recreated the original scenario, the text is considered successful and of "good quality." Results were analyzed using coding and aggregate scoring techniques, which are a form of quantitative research.

Compared to "Popularization Writing," the authors of "Direct Function" explain their methodology more effectively. They explain that statistical measures including text length (number of words) and lexical diversity (wide range of words) were used to indicate text quality in previous studies. Another common measurement of text quality involves the reader giving feedback, including via rubrics or holistic measurements (i.e., clarity, style, structure) (207-8). The type of reader impacts the feedback, and "raters" of a text tend to hone in on rubric guidelines and risk losing sight of judging the text as a whole, including its overall goal (208). It's worth noting that the "Popularization Writing" study uses these methods. Grabowski and Mathiebe argue that such measures are merely descriptive and not sufficiently inferential on their own (206-7). If writing is a "problem-solving process," then it's reasonable to assess text quality according to the reader's needs. Does the text say? What does the text *need* to say? How will the reader interpret the text? In short, the audience must not be ignored as a measurement of text quality. Once again, "Popularization Writing" did not include the audience in their study. Grabowski and Mathiebe propose the following **research questions**:

1. Does the quality of reporting texts, as measured through the readers' reconstructions of the described accident situation, systematically vary according to student writers' grade, school type, and family language? If the developed measure is valid and meaningful, it

should, by and large, repeat the general result patterns known from other assessment approaches.

2. How does the functional measure correlate with other approaches to text-quality assessment? Does it simply capture the same aspects, or does it reveal some unique variance of text quality?

The research questions are appropriate to test the authors' claims that text quality is determined by reader comprehension. The first question's purpose is to ensure that the authors' proposed assessment approach (reader comprehension) is verifiably robust and measurable. The second question expands on the first, not only identifying how reader comprehension overlaps with other assessment approaches, but how it also reveals new insights into textual analysis.

In the reference study, 277 German students from grades 5-9 across three different schooling levels were provided a picture of an incident, instructed to imagine the scenario taking place in the picture, and told to write an accident report for the police based on their interpretation. They wrote their reports by hand in one 10 minute class session. The resulting texts were edited for legibility and spelling errors so as not to affect text quality assessment (210).

In the current study, 277 university students participated in a 5-10 completely voluntary electronic assessment. Each student read one of the secondary school student's reports and were tasked to reconstruct what the accident looked like on a computer program (212-3). Participants of the current study were ignorant to the reference study's "writing instruction" and the original picture of the incident (213). Assigned text was available to reference throughout the 5 mins allotted for scene reconstruction. 277 individual recreations of the accident were generated as

evidence for analysis. If the recreations did not match the original picture, then the accident report describing the incident was not successful and, therefore, not "good quality."

Researchers coded and rated each reconstruction according to binary (1 = yes, 2 = no) codes in "three subscales with a total of 16 aspects, plus one holistic overall impression" (214). Thorough coding methods allowed them to determine whether each text was "sufficiently" successful at getting readers to understand (216). In other words, they made reader comprehension quantifiable. This evidence allowed authors to develop a *functionality score* (how similar is the original event to the reconstruction?) to assign to individual texts as a numerical indication of quality, which discourages subjective influence. A high score means a text's writing was clear enough for the reader to reimagine it accurately. The authors argue that report writing benefits the most from such scoring methods because the purpose of this genre is clear and measurable (217). However, it's unclear if this same functionality score would translate across genres, which is an opportunity for further research.

The authors conclude that their functionality score is a reliable and unique method for determining text quality (218, 223). Not only does it replicate results from other popular measurement methods, but it measures something different than the others that is overlooked: how well meaning is transferred from the writer to the reader. While traditional methods can determine if a text is *well written*, Grabowski and Mathiebe's functionality score determines how well a text *works* based on the genre. However, the authors point out that their methodology is not economical for researchers because it requires examination from an unbiased reader for each text (226).

Students typically aren't sure if their writing is "functionally successful" because they tend to receive critiques from teachers on "surface-level" issues rather than feedback from

unbiased readers with unique needs. Using their evidence, the authors conclude that collaboration is more effective for learning to write than being solely corrected by the teacher from rubrics. More specifically, students should collaborate with peers so that they can "check each other's text for impact and accuracy," give feedback to each other, and practice measuring text quality based on comprehension (224-5). The more a student practices giving and receiving feedback, the more they will inherently understand how to improve their texts by tailoring it to the reader (225).

This study could have also been done in a more qualitative, interview-based manner. Readers could be assigned to texts and asked to give feedback on the texts according to their comprehension. However, this approach is much harder to measure compared to coding and creating a functionality score, which would fundamentally change how they analyze their evidence. According to the authors, their goal was to create a new "methodological approach to the indication of text quality," and their coding and functionality scoring system is likely the most quantitative way to do so (222). As for recreating this study, changing the text from a report to a different genre would also fundamentally change the study. The authors agree that their scoring system is most suitable for informative texts rather than argumentative ones, which opens up opportunities for other researchers to study if reader comprehension can be similarly measured in different genres (224).

Where "Popularization Writing" performs a longitudinal, qualitative case study that analyzes how writing skills develop, "Direct Function" establishes a quantitative measurement to specifically measure how to determine text quality (i.e., how it performs). While they both use interviews, writing tasks, and coding to answer their research questions, Sterk et al.'s approach prompts them to ask more "why" questions ("Why aren't students competent at popularization

discourse after three years of undergraduate study?) while Grabowski and Mathiebe ask more questions about “how well” something functions (“To what extent can text quality be determined by the audience?”).

These articles taught me that doing research on rhetoric and writing are much more analytical and quantitative than I assumed. I am used to reading theory-based papers as opposed to research articles with specific methodologies. I admit, it was difficult to parse through a lot of the evidence in these pieces, but I have become much more competent at identifying the elements that make up an effective research article, including jargon and terminology associated with analyzing evidence (i.e., “scale aggregation,” “coding,” and "interrater agreement”). I’ve never been too interested in conducting research in this manner because it intimidated me; I’m still intimidated, but less so now that I’ve spent countless days reading dense research papers, even if I don’t entirely understand them.

Works Cited

Grabowski, J., & Mathiebe, M. (2024). “A Direct Functional Measure of Text Quality: Did the Reader Understand?” *Written Communication*, *41*(2), 203-229. https://doi.org/10.1177/07410883231222952. 2024.

Marnel Sterk, F., van Goch, M., Burke, M., & van der Tuin, I. (2025). “Popularization Writing Skills Development: A Longitudinal Case Study of the Writing Process and Writing Outcomes in Nine Undergraduate Interdisciplinary Students.” *Written Communication*, *42*(4), 966-998. https://doi.org/10.1177/07410883251349204. 2025.