

Measuring Writing

Dr. Niharika Rawat

Associate Professor,

N M Virani Science College, Rajkot

To understand both writing processes and writing interventions, as well as writing interventions themselves, it is sometimes necessary to assess text quality. However, the idea of 'text quality' is commonly misunderstood or misunderstood. A more precise understanding and explanation of what it means to be a good writer is the purpose of this book, according to the editors. Chapters in the book are broken down into the following categories: definitions of writing skill, rater effects, and automated essay scoring, to name a few.

Talent for writing can be interpreted in many ways.

This book begins with a discussion of whether writing skills may be applied to a variety of topic areas. Also, how many of a student's texts must be chosen before a generalisation about their writing skills can be drawn? The validity and generalizability of Schoonen's findings are demonstrated by the use of data collected from a large sample of students at three distinct times in time and for three separate writing projects in two languages. A student's written English as a Foreign Language competency may be gauged by looking at three to four tasks, each of which is graded by two raters, as explained by Schoonen for secondary school students. In L1, on the other hand, it is absolutely necessary. Seven to ten tasks with a double-rated grade are required.

The results show that generalizability in L1 tasks is significantly hampered because intra-writer variability in L1 activities is larger than in L2 writing. It's sad to say but that's the case. Findings show that tasks are more valuable to generalizability than raters are. Even among older students, writing in a foreign language is more stable than writing in one's own language, according to Van den Bergh et al results. 's (1st year university students vs. 9th

graders). Analysis on four dimensions is used to contrast the four dimensions of content (argumentation), conclusion (conclusion), and text structure. If you're looking for something that can be used across a wide range of subjects, you'll want to use analytical ratings rather than holistic assessments. Concerned about Schoonen's excessive quantity of tasks, Van den Bergh and colleagues are more reasonable: We conclude that using a single text as the only criterion for research yields no useful results. A reliable conclusion cannot be obtained from research that uses a single text as a criteria for evaluation. This is on page 32 of the book. The number of raters is the same. Any number of raters, ideally many, must come to an agreement on the overall quality of the various texts.

Methods of evaluation

In this part, we'll look at three distinct kinds of rating methods. To highlight the benefits and drawbacks of various rating techniques, Neumann uses multiple assessment studies based on research findings from the United States and Germany. These include holistic, analytical, and mixed models. Olinhouse, Santangelo, and Wilson continue to explain the American approach of evaluating writing. Most tests in the United States are only given once, are restricted to a specific genre, and are evaluated as a whole. Writing standards are seen as a degree of competency in these assessments, on the other hand (which comprises more genres and skills). Writing in L1 isn't a common practise, as the previous chapters have shown. It is also possible that results made from a single event and one genre that have been holistically scored may be confined to the genre under consideration, if at all This results in a picture of pupils' writing talents that is too narrowly focused. A wider range of tasks must be assessed, hence new solutions must be developed.

Input from a rater

The usage of a range of genres might lead to an increase in the variability of writing projects in terms of interpersonal relationships. Raters, on the other hand, might be a source of uncertainty (see also He et al. 2013). In the third segment, two chapters are devoted to this topic. More than 60 (!) different parameters were considered by Barkaoui and Knouzi's team of six raters as they evaluated two jobs based on around 20 attributes. Their rationale is that writing studies should take into account more than just test results. Despite the fact that this technique is exciting from a research aspect, it is not very practical or beneficial in

educational settings. While completing current writing assignments, Weigle and Montee focus on the use of textual borrowing as an important strategy. It's critical that students learn to read and write together since this reflects the kind of writing they'll encounter in college and the workplace. A lack of consensus among the five raters on how best to include sources into the essays was exhibited. There were also considerable variances in perspectives, which demonstrated the need of providing clear instructions and training to raters, particularly when they originate from varied backgrounds..

A computer software evaluates the content of an essay.

The book ends with the answer, which the reader may have already figured out by this time in the novel. Although it may appear that some processes cannot be automated, this is not necessarily the case. Ratings would be a lot easier if this were implemented, at the very least. What if some of the measures you're looking at are "simple to define" yet show strong relationships with more "complicated" measurements? Automated essay scoring is the subject of two chapters in this book. IntelliMetric and E-rater (both of which are American) are included in McCurry's overview of the Educational Testing Service's computer scoring tools (ETS). According to McCurry's review, machines can't give the same degree of grading validity that humans can. Withaus focuses on the problems of employing Automated Essay Evaluation for the analysis of multimodal writing in his last chapter. As McCurry did before, Withaus (together with Weigle and Montee) focuses on modern multimodal writing in addition to the strategies already mentioned in the chapter. There are no immediate solutions to this problem, however there are several possibilities for future research (also Burstein, Tetreault, & Chodorow, 2013; Deane, 2013). I recommend Chermis and Burstein's Handbook of AEE: Current Applications and New Directions if you want to learn more about Automated Essay Evaluation.

Here's what I think:

Must admit that the primary reason for my review was to learn about how to do a study on multilingual essay writing processes and how to enhance my own research by using the metrics in this book. In light of this new knowledge, I now know that I should employ activities that are suited for a variety of situations and genres, have been double-reviewed, and are evaluated holistically and analytically. However, did finishing the book answer my

initial question? Can you, on the other hand, point a fresh PhD student in the right direction for the answer? No, I can't help you with either of these requests. The book may be valuable to those who are looking for fine-grained information on scoring system characteristics (including automated techniques), processes, and the influence of raters. An important link is also made between the book's topics, first by the sharp and well-argued introduction, and subsequently by the individual authors. There are several cross-references and examinations of findings that contradict one another, as well as some answers in certain situations. Yet another flaw in this work is how exhaustively it covers every little aspect. The ability to pick an assessment strategy that is most suited to our research goals is something I, and I assume other researchers, would want to have in place. It is my wish that I could pick a number of texts, raters and so on that would match our requirements.

Many people who read it will reflect on the fact that they will never be able to do what they have set their minds to after completing. A minimum of three to four texts in a foreign language and seven to ten (!) written in the native tongue are required to be eligible for consideration. The texts will be evaluated by at least four raters on at least forty criteria. It's important to note that there's still a long way to go before we can begin to define text quality, particularly in terms of how to conduct reliable and practical research on the link between writing processes and text quality. Because these studies focus on more basic and methodological issues, it's probable that their conclusions won't be applicable to other, more practically oriented research projects.

Though the book makes it clear that having a small group of young students write one text in their mother tongue and then having this written work rated by a single rater is not a very sensible way to determine their writing competence, I believe that this rating is not a reliable indicator of their writing competence. According to the prologue of the book, "[...] writing skill cannot be reliably measured by means of a single writing product per writer," There must be many texts written by each author, as well as multiple raters for each literary output. Training raters, creating exact rating rubrics, and including benchmarks are all important concerns as well. the tenth page) If you want an accurate picture of a student's writing process in their native language, you need at least three processes in L2 and four processes (within a

genre) in L1 to acquire an accurate reading of their writing style (see also Rijlaarsdam et al., 2011).

It's possible that certain journals' limits on the dissemination of study findings contribute to the issue raised in the introduction, namely the fact that writing quality indicators are typically not well described to allow for comparisons between studies. Researchers may be required to be exceedingly concise when describing measurements due to the article's length limitation. This book provides a wealth of information for researchers working in the subject, and the sharing of questionnaires and statistical tools would be of great use to them. However, other media have become aware of this added value and now require detailed material descriptions to be supplied with their stories. Many other periodicals allow their readers to access additional content online. The Journal of Writing Research is one such example.

As previously said, I hope that scholars in the future will continue to produce relevant language dependent and independent measures that may be automated and made freely available to the scientific community.

References

- Burstein, J., Tetreault, J., & Chodorow, M. (2013). Holistic Discourse Coherence Annotation for Noisy Essay Writing. *Dialogue & Discourse*, 4(2), 34-52.
- Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18(1), 7-24.
- He, Tung-hsien, Wen Johnny Gou, Ya-chenChien, I. Shan Jenny Chen, and Shan-mao Chang. 2013. Multi-Faceted Rasch Measurement and Bias Patterns in EFL Writing Performance Assessment. *Psychological Reports* 112(2):469-485. doi: 10.2466/03.11.PR0.112.2.469-485.
- Shermis, M.D., & Burstein, J. (Eds) (2013). *Handbook of Automated Essay Evaluation: Current applications and new directions*. [Book review by Marie Stevenson]. *Journal of Writing Research*, 5(2), 239-243.
- Rijlaarsdam, G., Van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Van Steendam, E., Raedts, M. (2010). Writing. In S. Graham, A. Bus, S. Major, & L. Swanson (Eds.). *Application of Educational Psychology to Learning and Teaching*. APA Handbook. Volume 3 (p. 189-228). Washington, DC, US: American Psychological Association.