Title of Project:  
The role of voice in high-stakes second language writing assessment

Researcher:  
Cecilia Guanfang Zhao  
New York University  
gz312@nyu.edu

Research Supervisor:  
Dr. Lorena Llosa

Abstract:  
Although a construct commonly found in writing textbooks and rubrics, voice remains a concept that is only loosely defined in the literature and mystically assessed in practice. Few attempts have ever been made to formally investigate whether an authorial voice in a written text can be reliably measured, and how the strength of an authorial voice may affect the assessment of overall writing quality. Using a mixed-method approach, this study first developed and validated an analytic voice rubric (Phase 1), and then formally investigated the relationship between voice strength and writing quality (Phase 2) in the context of a high-stakes L2 writing assessment. Specifically, 400 TOEFL® iBT argumentative writing samples were used in the study with permission from Educational Testing Service, the copyright owner. Six voice raters, who had extensive experience in L1 and/or L2 writing instruction and assessment, rated the writing samples using a preliminary voice rubric developed on the basis of Hyland’s (2008) interactional model of voice. Each writing sample was double rated and the average of the two ratings was used in the subsequent statistical analysis. Qualitative data collected from rater think-aloud protocols and interviews were also analyzed to provide additional evidence on rubric reliability, validity and applicability.

Results from the first phase of the study showed that three categories in the preliminary voice rubric—personal aside, reference to shared knowledge, and the use of rhetorical and audience directed questions—were rarely identified in the TOEFL® iBT writing samples. They were thus excluded from all the subsequent statistical analyses. Factor analysis of the remaining voice elements pointed to a three-dimensional conceptualization of voice. The use of hedges, boosters and attitude markers together formed one dimension, which was later interpreted as a writer’s “manner of presentation.” In other words, a writer’s use of these linguistic devices indicates whether the author’s ideas are presented assertively, mildly, confidently, tentatively, enthusiastically, or maybe indifferently. A second dimension, consisting of authorial self-revelation and direct reader reference, was interpreted as “writer and reader presence.” Apparently, authorial self-revelation and direct reference to readers are also able to contribute to the realization of voice in written discourse. Finally, central point articulation and the use of directives together were considered as yet another component—“presence and clarity of ideas in the content.” In argumentative writing, particularly, the presence or absence of a central point also does make a difference in terms of getting a clear voice across.
Analyses of qualitative data, while confirming the three main dimensions identified in the quantitative analysis, revealed that the frequency-based preliminary voice rubric was limited in capturing the strength of an authorial voice. All raters agreed that while the presence and salience of those individual voice elements did influence their perception of voice strength in a writing sample, how these elements were used was probably more important in creating a strong authorial voice. Based on their think-aloud and interview data, therefore, a qualitative piece was added into the rubric. Thus, for each dimension, there were both frequency-based ratings for individual voice elements and a more qualitative evaluation of the general voice strength evoked by that dimension (e.g., voice strength evoked by the author’s manner of presentation).

This revised voice rubric was then used in the second phase of the study which examined the relationship between voice strength and writing quality. A series of correlation and regression analyses showed that overall voice strength was a significant predictor of L2 writing quality, and it was able to explain approximately 1/4 of the variance in the TOEFL writing scores ($R^2 = 0.246, F = 21.297, (df) = (3, 196), P <0.0001$). After controlling for the effect of other voice dimensions, however, only the content-related dimension (presence and clarity of ideas) was a statistically significant predictor of writing quality ($B = 0.39, t(196) = 4.42, p < 0.001$). Moreover, results from this phase also showed that a writer’s background had little, if any, influence on either the voice ratings or the relationship between voice and writing quality. Even after controlling for the effect of writer background variables, overall voice strength was still a significant and positive predictor of writing quality. Finally, moderation analysis showed that the relationship between voice and writing quality did not vary as a function of any particular writer background variables.

As the first formal attempt to develop and validate an analytic rubric that measures this well-known yet slippery construct of voice, this study offers an alternative three-dimensional conceptualization of voice. Unlike the few existing measures of voice that are mostly holistic and impression-based, this analytic rubric is more pedagogically useful in that it could help demystify the seemingly intangible concept of voice for both writing instructors and student writers. Using this rubric, teachers could make this notion of voice more accessible to their students, especially L2 students, and thus better help them to construct a strong and effective authorial voice with the proper use of those linguistic- and discourse-level language features.

The lack of a close alignment between what is proposed in Hyland’s (2008) theoretical model and what is observed in our empirical data, however, deserves careful interpretation. It should be noted that Hyland’s model was developed based on his examination of voice-related language features in a large corpus of published academic articles, whereas data used in this study were writing samples produced by TOEFL® iBT test takers within a limited time and in response to a single prompt. The difference in the nature of the writing, the length of the written products, as well as the writers’ level of language proficiency, thus, explains why certain more sophisticated voice elements (i.e., personal aside, reference to shared knowledge, and rhetorical or audience-directed questions) are rarely observed in our data.

What this suggests is two-fold. First, as much as we would want a general measure of voice that could be used across different assessment or educational settings, the evaluation of voice is probably not context-independent. Rather, we need to take into consideration the characteristics of the writing task, the genre and level of the writing, and the audience for whom the writing is intended. In different contexts, the realization of voice may be different; therefore, the criteria used to evaluate voice may vary. A second implication is that students’ ability to employ voice-related features in their writing probably depends on their general language
proficiency. Hence, a potentially useful way to address the concept of voice in writing assessment and instruction is to take a developmental approach. In other words, at different L2 proficiency levels, students could be taught to use different voicing strategies appropriate at that level, moving from the simplest and most straightforward ways of voicing, such as the use of first-person pronouns, gradually to more complex and sophisticated ways, such as inserting personal comments or reflections in the midst of a statement (i.e., personal aside).

Results on the relationship between voice and writing quality could also inform L2 writing pedagogy and contribute to the development of L2 writing assessment theory and practice in the long run. As some researchers have also argued (e.g., Connor & Mbaye, 2002), in addition to the linguistic aspect of writing that is traditionally focused on when evaluating L2 writing, the sociolinguistic and communicative aspects of writing also deserve our attention when assessing L2 writing competence. And the close association between voice and L2 writing quality found in this study offers one more reason for us to start looking into the communicative aspect of writing competence in L2 writing assessment.

The observation that, after controlling for the effect of other voice dimensions, the content-related dimension was the only significant, strong, and positive predictor of L2 argumentative writing quality also bears some important implications for L2 writing instruction. Stapleton (2002) argues that L2 writing researchers and practitioners should pay less attention to voice and “return the spotlight to ideas” (p 177). What Stapleton implies here is that voice and ideas are two separate constructs that are largely unrelated to each other. The present dissertation study, however, not only identifies, in empirical data, this dimension of ideas and content as an integral part of the concept of voice, but it also shows that it is indeed this dimension of voice that is most critically related to writing quality. As one of the voice raters also shared in the interview session, whether the content “makes sense” and the ideas “stand out” is indeed the sine qua non of a strong voice in writing, especially argumentative writing. Consequently, writing instructors should not teach voice in isolation, nor treat it merely as a matter of expression; rather, they should explicitly link idea and content development to the concept of voice for their students in their writing instruction.

Results from this study also showed that key writer background variables such as age, gender, L1 background had very little impact on the observed voice strength in L2 writing samples. What this suggests is that regardless of age, gender, language and cultural background, it is still possible for writers from different backgrounds to achieve a similar level of voice strength in their writing. It is also worth noting that in this study, most of the L2 writers with a non-Indo-European L1 background were from East Asian countries, overwhelmingly China, Japan, and Korea. Since these cultures are often identified as collectively oriented and vastly different from the Western culture of individualism, researchers have argued that L2 writers from these collectively-oriented cultural backgrounds often have difficulty voicing themselves in writing (e.g., Ramanathan & Atkinson, 1999; Ramanathan & Kaplan, 1996). Nonetheless, results from this study about the lack of association between voice strength and writer background variables seem to counter this argument. It might be possible, therefore, that with proper instruction, L2 writers could also write with a very strong voice, just like their native English-speaking peers.
References


Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. Assessing Writing, 10, 5-43


Leggo, C. (1991). Questions I need to ask before I advise my students to write in their own voices. Rhetoric Review, 10(1), 143-152.


