

Developing membership functions for scoring writing performance: A pilot study

Tan Jin, Barley Mak and Li Liu¹

¹Faculty of Education, The Chinese University of Hong Kong

tjin@cuhk.edu.hk, barleymak@cuhk.edu.hk, refilane@gmail.com

Abstract

This paper reports on an application of “membership functions” to score writing performance. The concept, membership functions, was originally used to deal with the problem of the indistinction between adjacent levels in scoring speaking performance (Jin, Mak, & Zhou, 2012; also see Jin & Mak, 2012a). Recently, the indistinction between adjacent levels has also been of concern in the holistic scoring of writing performance (see Liu, Mak, & Jin, 2012). In this connection, this paper applies membership functions to the holistic scoring of writing performance. The paper has three parts. First, new membership functions were constructed based on the IELTS Percentile Ranks 2010 (IELTS Website, 2 May 2012), which overall matched the TOEFL iBT Percentile Ranks 2010 (ETS Website, 2 May 2012a). Second, an example was provided to demonstrate how membership functions were used in the holistic scoring of writing performance. Finally, more empirical studies on the use of membership functions to score writing performance were called for in future research. This paper thereby contributes membership functions to the assessment of writing proficiency.

Keywords

membership functions, writing assessment, holistic scoring, writing performance, writing proficiency

1 Introduction

In scoring speaking tests, the indistinction is probably “inherent between adjacent levels” (Jin & Mak, 2012b, p. 15). To deal with the problem of the indistinction, the membership functions were applied to score speaking tests (Jin, 2009; Jin, Mak, & Zhou, 2012; Jin, Wang, Song, & Guo, 2008; also see Jin & Mak, 2012a). As for the scoring of writing tests, the indistinction between adjacent levels still exists (see Liu, Mak, & Jin, 2012). In using the holistic scoring of writing scripts, both

novice and experienced raters have reported the problem of fitting essays into a particular level (Erdosy, 2004; Sakyi, 2003). Therefore, this paper will develop membership functions for the holistic scoring of writing tests.

2 Constructing membership functions

The membership functions with five levels were constructed according to the method of Jin, Mak and Zhou (2012). The frequency distribution of IELTS 2010 academic candidates (IELTS Website, 2 May 2012) was used. It was also found that the IELTS Percentile Ranks 2010 (IELTS Website, 2 May 2012) overall matched the TOEFL iBT Percentile Ranks 2010 (ETS Website, 2 May 2012a) based on the report linking TOEFL iBT scores to IELTS scores (2 May 2012b).

Figure 1 is the membership functions constructed for the holistic scoring of the writing assessment, comprising five levels, namely Level 1, Level 2, Level 3, Level 4 and Level 5.

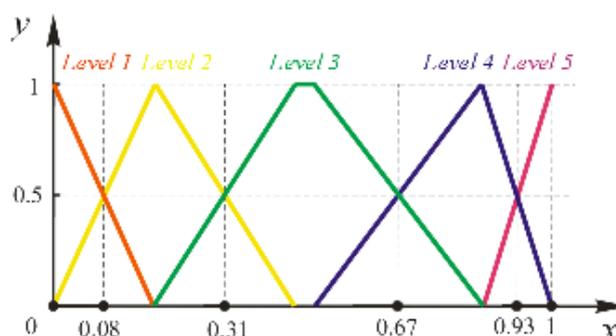


Figure 1: The Membership Functions

3 An example using the membership functions

In scoring writing performance, raters allocate a score from 1 to 10 to the two adjacent levels with a total of 10, for example, 2 and 8, 4 and 6; the raters can give a 10 to one particular level if they are very confident of the level (Jin, Mak, & Zhou, 2012).

Below is an example using the membership functions. Rater A scores a candidate's writing script using holistic scoring. Based on the holistic rating scale, Rater A assigns the scores as follows:

4 to Level 3; 6 to Level 4

The 4 and 6 are subsequently standardized as decimals (i.e., $4 \rightarrow 0.4$ and $6 \rightarrow 0.6$). For easy score use and interpretation, these two scores (0.4 and 0.6) have to be converted to a final score using the membership functions (Jin, Mak, & Zhou, 2012). The two standardized scores (i.e., 0.4 and 0.6) with their corresponding levels (i.e., Level 3 and Level 4) are demonstrated in Figure 2. The center of the shaded area is the final score. The Center of Gravity (COG) technique (Van Leekwijck & Kerre, 1999) is thus used to compute the value of the center—the final score. As can be seen in Figure 2, the value of the center of the shaded area is 0.63, meaning that the final score of this candidate is 0.63.

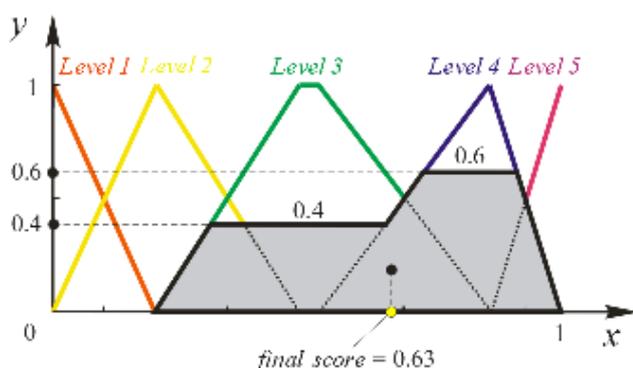


Figure 2: The Example Using the Membership Functions

4 Conclusion

In this paper, the membership functions were used to deal with the problem of the indistinction between adjacent levels in the assessment of writing proficiency. The membership functions were first constructed based on the frequency distribution of IELTS 2010 academic candidates (IELTS Website, 2 May 2012), which were also consistent with the TOEFL *iBT* Percentile Ranks 2010 (ETS Website, 2 May 2012a). Using the membership functions constructed, an example was then given to illustrate how the membership functions worked in the holistic scoring of writing tests. As the concept, membership functions, is comparatively new in the writing assessment field, it is suggested that more studies be conducted in future to validate the membership functions developed here.

References

- Erdosy, M. U. (2004). *Exploring variability in judging writing ability in a second language: A study of four experienced raters of ESL compositions* (TOEFL Research Report No. RR-03-17). Princeton, NJ: Educational Testing Service.
- ETS Website (2 May 2012a). *Percentile Ranks for TOEFL iBT Scores (Total Group)*. Retrieved on 2 May 2012 from <http://www.ets.org/Media/Research/pdf/TOEFL-SUM-2010.pdf>
- ETS Website (2 May 2012b). *Linking TOEFL iBT™ Scores to IELTS® Scores—A Research Report*. Retrieved on 2 May 2012 from http://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf
- IELTS Website (2 May 2012). *IELTS Percentile ranks 2010 (Academic Candidates)*. Retrieved on 2 May 2012 from http://www.ielts.org/researchers/analysis_of_test_data/percentile_ranks_2010.aspx
- Jin, T. (2009). *An investigation into fuzzy scoring methods for speaking tests* (master thesis). Shanghai: Shanghai Jiao Tong University, Beijing: China Master Theses Full-text Database.
- Jin, T., & Mak, B. (2012a). Confidence scoring [Audio podcast]. *Language Testing Bytes*, 8 (Podcasts for *Language Testing*).
- Jin, T., & Mak, B. (2012b). Distinguishing features in scoring L2 Chinese speaking performance: How do they work? *Language Testing*. Advance online publication (21 May 2012). doi: 10.1177/0265532212442637.
- Jin, T., Mak, B., & Zhou, P. (2012). Confidence scoring of speaking performance: How does fuzziness become exact? *Language Testing*, 29(1): 43-65.
- Jin, T., Wang, Y., Song, C., & Guo, S. (2008). An empirical study of fuzzy scoring methods for speaking tests. *Modern Foreign Languages*, 31(2): 157-164.
- Liu, L., Mak, B., & Jin, T. (2012). Measuring content quality of writing performance: Putting content in its place. *The 34th Language Testing Research Colloquium* (LTRC 2012), Princeton, New Jersey, USA.
- Sakyi, A. A. (2003). *A study of the holistic scoring behaviors of experienced and novice ESL instructors*. (Unpublished doctoral dissertation). Toronto: University of Toronto.
- Van Leekwijck, W., & Kerre, E. E. (1999). Defuzzification: Criteria and classification. *Fuzzy Sets and Systems*, 108(2), 159-178.