Appendix

Rater Training Process for Writing Section

I.   Introduction

- Introduce the purpose and format of the Proulex EPT B1 writing sections
- Introduce the scoring rubrics for writing to rater candidates with benchmark (anchor) writing samples which typify four levels in four categories (integration, task achievement, organization, and language use) for reading-to-write task and in three categories (task achievement, organization, and language use) for extended writing task – you need at least 8 benchmark writing samples (four for reading-to-write and another four for extended writing tasks)

II.   Practice

- Provide rater candidates with a set of pre-rated writing samples (4 for each writing task) and ask them to score each of them independently
- Discuss rater candidates' ratings on 8 writing samples, and then present the benchmark analytic ratings for each sample

III.   Certification

- Provide rater candidates with 16 writing samples (8 for each writing task), and ask them to score each sample independently without discussion
- Ask them to check their ratings against the benchmark ratings
- Discuss their ratings if there is any discrepancy between their ratings and benchmark ratings
- Identify rater candidates whose agreement rate is below 70%, who should not be qualified to be rater. Provide them additional samples (4 for each task) to see if their agreement rate is higher than the threshold level (70%)

N.B. It is extremely important to keep track of raters' scoring in both training and actual test scorings. Enter all the data each rater candidate rate on each category for each writing sample during the training session in the Excel spreadsheet. Once they are certified, their actual ratings need to be stored and analyzed for ensuring their inter-and intra-rater reliabilities.

Zepeda Orozco, R. A. & Shin, S-Y. (2019) Developing and Validating an English Proficiency Test. *MEXTESOL Journal*, *43*(3).