



Title of Project:

Assessing Speakers of World Englishes: The Roles of
Rater Language Background, Language Attitude and Training

Researcher:

Jing Wei
New York University
weijingw@gmail.com

Research Supervisor:

Lorena Llosa
New York University
lorena.llosa@nyu.edu



Jing Wei

Project Summary

Importance of the topic

The purpose of this dissertation is to explore the feasibility of assessing alternative constructs other than Inner Circle English varieties in standardized international language tests. As English is spreading globally, new linguistic features that deviate from the Inner Circle English forms have started to emerge. Although the features characterizing the target language use domain have changed, English language tests still remain unchanged, which causes a mismatch between the forms that are tested and the forms that are required for successful communications in the real world. Tests that continue to target Inner Circle English varieties are biased against users of other English varieties, as their test scores do not reflect their true communicative competence.

Given the mismatch between what is tested and what is required for real world communications, World Englishes (WE) scholars (e.g., Elder and Davies, 2006) called for a revolution in language tests so that constructs other than native English varieties can be assessed. They argued that instead of using Inner Circle native English varieties as the standard, language tests should be normed on Outer and Expanding Circle English varieties (traditionally labeled as non-native English). On the other hand, researchers on English as a Lingua Franca (ELF) (e.g., Jenkins, 2006; Jenkins & Leung, 2013, 2017) went even further by contending that English use has transcended the boundaries amongst language varieties. English language users are constantly shuttling between different varieties of English or even between different languages. Therefore, knowledge/ability of using one single variety or language is not sufficient to meet the communicative demands in the current world. A language user needs to be proficient in multiple languages and language varieties as well as to have the ability to make the appropriate choice about which language or variety to use based on a communicative context. ELF scholars (Jenkins & Leung, 2013) propose a two-step approach to develop the assessment of ELF. The preliminary step is to not penalize emergent variants of ELF in current language tests. The next step is to overhaul the entire tests and reconstruct the entire test based on a new model of assessment that



reflects the reality of ELF communications. The purpose of this dissertation is to take the preliminary step of moving towards the assessment of ELF by exploring the possibility of training raters to accept Chinese and Indian English varieties in international language tests such as TOEFL iBT.

Research questions

This study was guided by the following three research questions:

- 1) To what extent does a shared L1 effect exist when Chinese and Indian raters are rating Chinese and Indian test takers' responses to TOEFL iBT Speaking Tasks?
- 2) How does a special training program that raises raters' awareness of the unique features characterizing Chinese and Indian varieties of English affect raters' scores and their awareness of those features?
- 3) How do raters' attitudes towards Outer and Expanding Circle varieties of English relate to their scores?

Methods

This study employed a mixed methods design with an experiment component to examine the above three research questions. Indian, Chinese and American raters were first asked to score Indian and Chinese test taker TOEFL iBT speaking responses. Next, they were randomly assigned into a group that used the special training package and a control group that was trained with the regular training package. In order to measure the effects of training on their scores and awareness of World Englishes features, both the special and regular training groups scored a set of speaking responses after the training. They also performed think-aloud protocols (TAP) on a subset of responses to demonstrate the decision-making process and the criteria they applied in assigning scores. Finally, all raters completed a survey questionnaire that targets their attitudes towards English varieties and their views on issues related to English as a Lingua Franca.

Thirty Indian, Chinese, and American raters (i.e., 10 raters per language group) participated in this study. Almost all of the raters were graduate students from New York University and Columbia University with background in TESOL, English education or a related field. Raters' ages ranged from 21 to 63, with a mean of 29.29. Among the raters, 7 were male and 23 were female. Each language group has a mix of experienced and novice teachers, whose teaching experience ranged from 0 to 6 years.

Findings

To address Research Question 1 that examines whether a shared L1 advantage in scoring existed, the FACET results show that shared L1 advantage only existed with Chinese raters' scores of Chinese test takers but not with Indian raters' scores of Indian test takers. The interactional analyses of rater groups' scores of test taker groups indicate that Chinese test takers' scores were increased by 0.09 on a scale of 1-4 when rated by Chinese raters than by the combined group of American and Indian raters.

To address Research Question 2-- how does a special training package that target Outer and Expanding Circle varieties of English affect raters' scores, the FACET results show that there was more variation in raters' scores after they received the special training than after they received the regular training. The analyses revealed that raters who were normed on features



unique to Chinese and Indian English varieties scored more consistently internally but had larger spread in severity measures compared to those who were normed on standard American English.

The analyses of TAP data provided a possible explanation of why rater severity spread was larger for the special training group than for the regular training group: raters within a training group showed different degrees of acceptance towards features of Chinese and Indian varieties of English. In other words, raters became more aware of what features are unique to Chinese and Indian Englishes after they received the special training, but not all of them considered it as legitimate to use those features in the context of a high-stakes standardized test.

To address Research Question 3—how raters’ attitude towards Outer and Expanding Circle varieties of English relates to their scores, analyses of rater language attitude questionnaire responses show that for speech samples with higher scores, there was no variation in raters’ semantic ratings about qualities of the speakers. However, for speech samples with lower scores, raters differed from each other in their judgment about the educational level, intelligence and confidence level of the speakers.

Implications

The current study has significant theoretical, methodological, and practical implications for the field of language assessment. Theoretically, the result of the study has implications for the assessment of English as a Lingua Franca. Raters have shown some awareness about the global spread of English and its impact on language assessment. However, not all raters have had a clean break from the native speaker ideology, even after being educated about features of Outer and Expanding Circle English varieties. It is not possible to develop English as a Lingua Franca test unless test raters and test users are ready to embrace the linguistic and cultural diversity in the target language use domain and the need for that reality being reflected in the assessment practice. As for its implications for the rater training practice, the findings from this study indicate that rater training should expand its focus from educating raters about language features to incorporating a component that targets raters’ language attitude. Even after becoming aware of language features unique to an English variety, raters that are dominated by native speaker ideology may still consider those features as inappropriate for the context of a standardized test.

Methodologically, this study shows the importance of employing a mixed-method design to investigate training effects both in terms of their scores and scoring criteria. The analyses of scores showed that raters from the special training group had higher severity spread than raters from the regular training group. Such findings about rater scores can be supported by TAP data, which showed that not all raters who received the special training accepted Chinese and Indian English variety features in spite of being aware of them. The triangulation of both qualitative and quantitative data provided insights that would not be obtained, if this study only focused on quantitative analyses of rater scores.



References

- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Ajzen, I., & Timko, C. (1986). Correspondence between health attitudes and behavior. *Basic and Applied Social Psychology*, 7(4), 259-276.
- Ashcroft, B., Griffiths, G., & Tiffin, H. (1989). *The empire writes back: Theory and practice in post-colonial literature*. London, UK: Routledge.
- Barnwell, D. (1989). 'Native' native speaker and judgments of oral proficiency in Spanish. *Language Testing*, 6, 152-163.
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustic Society of America*, 114(3), 1600-1610.
- Brennan, E., & Brennan, J. (1981). Accent scaling and language attitudes: Reactions to Mexican-American English speech. *Language and Speech*, 24, 207-221.
- British Council, IELTS Australia, University of Cambridge ESOL Examination. (2012). IELTS Speaking band descriptors [Assessment instrument, Public Version]. Location: British Council. Retrieved from https://takeielts.britishcouncil.org/sites/default/files/IELTS_Speaking_band_descriptors.pdf
- Brown, K. (1968). Intelligibility. In A. Davies (Ed.), *Language testing symposium* (pp. 180-191). Oxford, UK: Oxford University Press.
- Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing*, 12, 1-15.
- Brown, J. D. (2014). The future of World Englishes in language testing. *Language Assessment Quarterly*, 11, 5-26.
- Brutt-Griffler, J., & Samimy, K. K. (2001). Transcending the nativeness paradigm. *World Englishes*, 20(1), 99-106.
- Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly*, 3, 229-242.
- Canagarajah, S. (2007). Lingua Franca English, Multilingual Communities, and Language Acquisition. *The Modern Language Journal*, 91, 923-939.



- Carey, M. D., Mannell, R. H., & Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, 28(2), 201-219. doi:10.1177/0265532210393704
- Cargile, A. C., Giles, H., Ryan, E. B., & Y. Bradac, J. J. (1994). Language attitudes as a social process: A conceptual model and new directions. *Language and Communication*, 14, 211-236.
- Chalhoub-Deville, M., & Wigglesworth, G. (2005). Rater judgment and English language speaking proficiency. *World Englishes*, 24, 383-391.
- Connor-Linton, J. (1995b). Cross-cultural comparison of writing standards: American ESL and Japanese EFL. *World Englishes*, 14, 99-115.
- Crowley, T. (1996). *Language in history: Theories and texts*. London, UK: Routledge.
- Crystal, D. (1997). *English as a global language*. Cambridge, UK: Cambridge University Press.
- Crystal, D. (2004). *The language revolution*. Cambridge, UK: Polity.
- Csizér, K., & Kontra, E. H. (2012). ELF, ESP, ENL and their effect on students' aims and beliefs: A structural equation model. *System*, 40, 1-10.
- Davidson, F. (2006). World Englishes and test construction. In B. B. Kachru, Y. Kachru, & C. L. Nelson (Eds.), *The handbook of world Englishes* (pp. 709-717). Malden, MA: Wiley – Blackwell.
- Davies, A. (1989). Is international English an interlanguage? *TESOL Quarterly*, 23(3), 447-467.
- Davies, A. (1999). Standard English: Discordant voices. *World Englishes*, 18, 171-186.
- Davies, A., Hamp-Lyons, L., & Kemp, C. (2003). Whose norms? International proficiency tests in English. *World Englishes*, 22(4), 571-584.
- De Vaus, D. (2002). *Analyzing social science data: 50 key problems in data analysis*. Los Angeles, CA: Sage.
- Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics*, 26, 282-301.
- Elder, C., & Harding, L. (2008). Language testing and English as an international language: Constraints and contributions. *Australian Review of Applied Linguistics*, 31, 34, 1-34.
- Elder, C., Knoch, U., Barkhuizen, G., & Randow, J. (2005). Individual feedback to enhance rater training: Does it work? *Language Assessment Quarterly*, 2(3), 175-196.



- Elder, C., Knoch, U., Barkhuizen, G., & Randow, J. (2007). Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing*, 24(1), 37-64.
- Fairclough, N. (Ed.) (1992). *Critical Language Awareness*. London, UK: Longman.
- Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37, 313-26.
- Fazio, R. H. (1986). How do attitudes guide behaviors? In R. M. Sorrention & E. T. Higgins (Eds.), *The handbook of motivation and cognition: Foundations of social behavior* (Vol. 1) (pp. 204-243). New York, NY: Guilford Press.
- Gao, X. (2012). The study of English as a patriotic enterprise. *World Englishes*, 31(3), 351-365.
- Graddol, D. (1997). *The future of English?: A guide to forecasting the popularity of the English language in the 21st century*. London, UK: British Council.
- Graddol, D. (2001). The decline of the native speaker. *AILA Review*, 13, 57-68. Retrieved from <http://www.aila.info/en/publications/aila-review/review-volumes/75-aila-review-issue-13.html>
- Gnutzmann, C. (2000). Lingua franca. In M. Byram (Ed.), *The Routledge encyclopedia of language teaching and learning* (pp. 356-359). London, UK: Routledge.
- Hamp-Lyons, L., & Zhang, B. W. (2001). World Englishes: Issues in and from academic writing assessment. In L. Flowerdew & M. Peacock (Eds.), *Research perspective on English for academic purposes* (pp. 101-16). Cambridge, UK: Cambridge University Press.
- Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing*, 29(2), 163-180.
- Hill, C., & Parry, K. (Eds.). (1994). *From testing to assessment: English as an international language*. London, UK: Longman.
- Hill, K. (1997). Who should be the judge? The use of non-native speakers as raters on a test of English as an international language. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Eds.), *Current developments and alternatives in language assessment: Proceedings of LTRC 96* (pp. 275-290). Jyvaskyla, Finland: University of Jyvaskyla and University of Tampere.
- Hinkel, E. (1994). Native and nonnative speakers' pragmatic interpretations of English texts. *TESOL Quarterly*, 28, 353-376.



- Hrubes, D., Ajzen, I., & Daigle, J. (2001). Predicting hunting intentions and behavior: An application of the theory of planned behavior. *Leisure Sciences*, 23(3), 165-178.
- Hughes, A., & Lascaratou, C. (1982). Competing criteria for error gravity. *ELT Journal*, 26, 175-82.
- Hsu, H. (2012). *The impact of World Englishes on oral proficiency assessment: Rater attitude, rating tendency, and challenges* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Champaign, IL.
- Jenkins, J. (2009). English as a lingua franca: interpretations and attitudes. *World Englishes*, 28(2), 200–207.
- Johnson, K., & Johnson, H. (Eds.). (1998). *The Encyclopedic dictionary of applied linguistics*. Oxford, UK: Blackwell.
- Kachru, B. B. (1982). *The other tongue: English across cultures*. Urbana, IL: University of Illinois Press.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge, UK: Cambridge University Press.
- Kang, O. (2008). *Ratings of L2 oral performance in English: Relative impact of rater characteristics and acoustic measures of accentedness* (Unpublished doctoral dissertation). University of Georgia, Athens, GA.
- Kim, H. (2005). *World Englishes and language testing: The influence of rater variability in the assessment process of English language oral proficiency* (Unpublished doctoral dissertation). University of Iowa at Iowa City, IA.
- Kim, Y. (2009). An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing*, 26, 187-217.
- Kobayashi, T. (1992). Native and nonnative reactions to ESL compositions. *TESOL Quarterly*, 26, 81-112.
- Kobayashi, H., & Rinnert, C. (1996). Factors affecting composition evaluation in an EFL context: Cultural rhetorical pattern and readers' background. *Language Learning*, 46, 397-437.
- Kondo, Y. (2010). Examination of rater training effect and rater eligibility in L2 performance assessment. *Pan-Pacific Association of Applied Linguistics*, 14(2), 1-23.



- Knoch, U., Read, J., Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing writing*, 12, 16-43.
- Kramarae, C., & Treichler, P. A. (1985). *A feminist dictionary*. Boston, MA: Pandora.
- Lazaraton, A. (2005, May). *Non-native speakers as language assessors: Recent research and implications for assessment practice*. Paper presented at the ALTE, Berlin, Germany.
- Lindermann, S. (2002). Listening with attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*, 31(3), 419-441.
- Lindermann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics*, 7(3), 348-364.
- Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. New York, NY: Routledge.
- Lowenberg, P. H. (1993). Issues of validity in tests of English as a world language: Whose standards? *World Englishes*, 12(1), 95-106.
- Lowenberg, P. H. (2000). Assessing English proficiency in the global context: The significance of non-native norms. In H. W. Kam (Ed.), *Language in the global context: Implications for the language classroom* (pp. 207-228). Singapore: SEAMEO Regional Language Center.
- Lowenberg, P. H. (2002). Assessing English proficiency in the expanding circle. *World Englishes*, 21, 431-435.
- Lukmani, Y. (2002). *English in India: Assessment issues*. Presentation at the Hong Kong Seminar.
- Lumley, T., & McNamara, T. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, 12, 54-71.
- Mackey, L. S., & Finn, P. (1997). Effect of speech accent on speech naturalness ratings: A systematic replication of Martin, Haroldson, and Triden (1984). *Journal of Speech, Language & Hearing Research*, 40, 349-361.
- Major, R. C., Fitzmarice, S., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, 36(2), 173-190.
- McArthur, T. (1987). The English languages? *English Today*, 11, 9-13.



- McNamara, T. (1996). *Measuring second language performance*. Essex, Harlow, UK: Pearson Education Limited.
- Milroy, J., & Milroy, L. (1985). *Authority in language*. London, UK: Routledge.
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111-131.
- Pihko, M. K. (1997). "His English sounded strange": *The intelligibility of native and non-native English pronunciation to Finnish learners of English*. Jyvaskyla, Finland: Centre for Applied Language Studies.
- Pulcini, V. (1997). Attitudes toward the spread of English in Italy. *World Englishes*, 16(1), 77-85.
- Rampton, B. (1990). Displacing the 'native speaker': Expertise, affiliation, and inheritance. *ELT Journal*, 44, 97-101.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511-531.
- Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of non-native English speaking teaching assistants. *International Journal of Intercultural Relations*, 14, 337-353.
- Santos, T. A. (1988). Professors' reactions to the academic writing of non-native speaking students. *TESOL Quarterly*, 22(1), 69-90.
- Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25(4), 553-581.
- Seidlhofer, B. Breteneder, A., & Pitzl, M. (2006). English as a lingua franca in Europe: Challenges for applied linguistics. *Annual Review of Applied Linguistics*, 26, 3-34.
- Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing*, 18, 303-325.
- Shohamy, E., Gordon, C., & Kraemer, R. (1992). The effects of raters' background and training on the reliability of direct writing tests. *The Modern Language Journal*, 76, 27-33.
- Stibbard, R. M., & Lee, J. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *Journal of the Acoustic Society of America*, 120(1), 433-442.



- Southworth, F. C. (1985). The social context of language standardization in India. In N. Wolfson & J. Manes (Eds.), *Language of inequality* (pp. 225-40). Berlin, Germany: Mouton.
- Swann, J., Dermert, A., Lillis, T., & Mesthrie, R. (2004). *A dictionary of sociolinguistics*. Retrieved from http://www.credoreference.com/entry/edinburghds/standard_language_variety
- Wei, J., & Llosa, L. (2015). Investigating differences between American and Indian raters in assessing TOEFL iBT Speaking Tasks. *Language Assessment Quarterly*, 12(3), 283-304.
- Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing*, 11, 197-223.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263-287.
- Widdowson, H. G. (1994). The ownership of English. *TESOL Quarterly*, 28(2), 377-389.
- Wilcox, G. K. (1978). The effect of accent on listening comprehension: A Singapore study. *English Language Teaching Journal*, 32, 118-127.
- Xi, X., & Mollaun, P. (2011). Using raters from India to score a large-scale speaking test. *Language Learning*, 61(4), 1222-1255. doi:10.1111/j.1467-9922.2011.00667.x
- Zahn, C. J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology*, 4, 113-124.
- Zhang, W. X. (1999). The rhetorical patterns found in Chinese EFL student writers' examination essays in English and the influence of these patterns on rater response. Unpublished PhD thesis, Hong Kong Polytechnic University.
- Zhang, Y., & Elder, C. (2010). Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? *Language Testing*, 28(1), 1-20.