

Speaker self-ratings in the field: A discussion of best practice

Researchers interested in language change in heritage or immigrant languages face a number of unique challenges when determining speaker proficiency. Standard practice in many areas of language research involves speaker self-assessment, despite research showing that self-ratings can be influenced by language anxiety and do not always accurately reflect linguistic capabilities (MacIntyre, Noels and Clément 1997). Although the bi-/multilingual reality is not captured through proficiency alone (Anderson et al. 2018), self-ratings provide a quick measure for speaker assessment and continue to be utilized in combination with formal testing tools in lab and field settings. However, heritage speakers may not be familiar with reflecting on and assessing their language skills, or they are not literate in all of their languages, making both self-ratings and established proficiency tests difficult to employ and interpret.

Self-ratings are arbitrary and often difficult to reconcile with actual speaker performance. For instance, in a comparison study of grammatical case in New Ulm German (NUG) and North American Icelandic (NAI), participants provided self-ratings of speaking proficiency that indicated one community was much more proficient than the other (NAI: 6.6/10; NUG: 7.5/10 (Schwarz 2019)). However, performance on a story-telling task showed the opposite effect. Despite giving themselves a lower score than the NUG speakers gave themselves, NAI speakers produced more structurally diverse sentences, see ex. (1a-b). Only 14% of sentences produced by NUG speakers included more than one argument, versus 84% of sentences produced by NAI speakers. By this measure, there is a clear discrepancy between the self-ratings and performance.

- (1) a. *hier d maedl ist mit d bub und d bub hat a ball* (NUG #101)
here the girl is with the boy and the boy has a ball
'here the girl is with the boy and the boy has a ball'
- b. *hún fær boltann frá drengnum* (NAI #GM04)
she get ball.indef from boy.def
'she gets a ball from the boy'

In this talk, we invite discussion about best practices for obtaining proficiency scores from heritage speakers in a manner that is time and cost efficient, but still constitutes an accurate reflection of the speakers' linguistic abilities. First and foremost we must define what it means to be proficient and what constitutes proficiency, in particular in the heritage language. We then present data from interviews that highlight speakers' difficulty and discomfort with self-assessment. We discuss how linguistic questionnaires and proficiency tests can be combined to provide a baseline for comparison between communities while avoiding raising the speakers' affective barrier. We touch on the advantages and disadvantages of various post-hoc proficiency measurements, in the vein of Petersen et al. (2018), calculate these scores for three heritage speaker communities in North America (New Ulm German, Plautdietsch and North American Icelandic speakers), and compare them with the respective speaker self-ratings.

References: (1) Anderson, John A. E., Mak, Lorinda, Keyvani Chahi, Aram & Bialystok, Ellen (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods* 50, 250-263. DOI 10.3758/s13428-017-0867-9. (2) MacIntyre, Peter D., Noels, Kimberly A., & Clément, Richard (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language learning* 47(2), 265-287. (3) Petersen, Jan H., Hansen, Gert F., Thøgersen, Jacob, & Kühl, Karoline (2018) Linguistic Proficiency: A Quantitative Approach to Immigrant and Heritage Speakers of Danish. *Corpus Linguistics and Linguistic Theory*. doi: <https://doi.org/10.1515/cilt-2017-0088> (4) Schwarz, Lara S. (2019). *(In)stability in heritage Germanic: Examining the role of form and function*. Penn State University dissertation.