Multilingual Language Diversity (MLD) Score calculated from LHQ3

The LHQ3 Development Team

Many colleagues have used the LHQ (Li, Sepanski, & Zhao, 2006) and LHQ 2.0 (Li, Zhang, Tsai, & Puls, 2014). Recently, we have introduced LHQ3 (Li, Zhang, Yu, & Zhao, 2019) with a new aggregated score function to represent participants' overall language background and usage information. The LHQ now also includes the Multilingual Language Diversity (MLD) score, which was not discussed in the published paper of Li et al. (2019). The MLD is inspired by recent work such as Gullifer and Titone (2019) and DeLuca, Rothman, Bialystok, and Pliatsikas (2019), according to which we can better describe bilingualism through language usage in terms of context and diversity. Specifically, following Gullifer and Titone (2018, 2019), we use the concept of Shannon Entropy (*H*) to describe such diversity, for example, through the estimated daily usage time of each language (Gullifer & Titone, 2019). The MLD calculates the H based on a participant's dominance in each language (rather than just daily usage time), and is now part of the aggregated scores automatically calculated by LHQ3. We encourage researchers who use the MLD score to cite Gullifer and Titone (2019), who developed and reported the novel insight of using entropy to characterize multilingual language diversity.

The MLD score is calculated as follows:

(1) For the i^{th} language reported by a participant, LHQ3 calculates an aggregated dominance score ($Dominance_i$) based on both the participant's self-reported proficiency and frequency of usage time (hours per day) on different components of the language (see Equation 3 of Li, Zhang, Yu & Zhao, 2019 for details). LHQ3 then calculates a temporary variable for the i^{th} language that we term as Proportion of Dominance (PD_i):

$$PD_i = \frac{Dominance_i}{\sum_{i=1}^{n} Dominance_i}$$

where n represents the total possible languages a participant has learned. PD_i roughly represents the proportion of a language$_i$ (i^{th} language) that is dominant in a participant's language environment/usage.

(2) The participant's overall multilingual language diversity (MLD) level is then calculated in a form of Shannon Entropy (*H*)

$$H = -\sum_{i=1}^{n} PD_i \log_2(PD_i)$$

H will be in a range between 0 and 2. As shown in Table 1 below, when a participant is monolingual (ID *ufvs2*), the PD_i for her solo language will be 1 and MLD generates a zero for her language diversity measurement. In another example (ID

vlfm9), *H* will be 1 if a participant is a bilingual with a more balanced usage and fluency of her two languages. H as a score of 1 is calculated as follows:

$$H = -(0.5 * \log_2 0.5 + 0.5 * \log_2 0.5) = -[0.5 * (-1) + 0.5 * (-1)] = 1$$

Since LHQ3 allows participants to self-report their LHQ data for up to 4 languages, *H* will be 2 in a most extreme case when a participant uses all 4 languages equally in terms of both proficiency and frequency: an extreme H score of 2 means that a multilingual speaker of 4 languages has the largest bilingual language diversity that LHQ3 can give (participant *zgbh6* would be a close example).

Table1: Example MLD scores from LHQ3 for several multilingual cases*

Participant ID	L1 Pro	L2 Pro	L3 Pro	L4 Pro	L1 Dom	L2 Dom	L3 Dom	L4 Dom	MLD Score
beqi9	1	0.75	0	0	0.67	0.46	0	0	**0.97**
fqch6	1	0.96	0.96	0	0.59	0.58	0.58	0	**1.58**
fulp3	1	0.25	0	0	0.67	0.16	0	0	**0.71**
ufvs2	1	0	0	0	0.8	0	0	0	**0**
vifm9	1	1	0	0	0.64	0.64	0	0	**1**
zgbh6	1	1	0.75	0.36	0.6	0.6	0.41	0.2	**1.89**

Note: * these are not real data from subjects; the numbers provided here are for illustration purposes. Pro = Proficiency score; Dom = Dominance score; MLD Score = Multilingual Language Diversity Score

Reference

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7565–7574.

Gullifer, J. W., & Titone, D. (2018). Compute language entropy with {languageEntropy}. Retrieved from https://github.com/jasongullifer/languageEntropy. http://doi.org/10.5281/zenodo.1403272

Gullifer, JW. & Titone, D. (2019). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition 1–12*.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaires: A web-based interface for bilingual research. *Behavior Research Methods*, *38,* 202-210.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language History Questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3)*,* 673680.

Li, P., Zhang, F., Yu, A., & Zhao, X. (2019). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition 1–7*.