Bilingualism: Language and Cognition

cambridge.org/bil

A Content Overlap Analysis of bilingualism questionnaires: Considering diversity

Ronessa Dass[1], Irina Smirnova-Godoy[2], Olivia McColl[2], John G. Grundy[3], Gigi Luk[4] and John A. E. Anderson[2]

[1]McMaster University, Hamilton, Ontario, Canada; [2]Carleton University, Ottawa, Ontario, Canada; [3]Iowa State University, Ames, IA, USA and [4]McGill University, Montreal, Quebec, Canada

Research Notes

Cite this article: Dass, R., Smirnova-Godoy, I., McColl, O., Grundy, J.G., Luk, G., & Anderson, J.A.E. (2024). A Content Overlap Analysis of bilingualism questionnaires: Considering diversity. *Bilingualism: Language and Cognition*, **27**, 744–750. https://doi.org/10.1017/S1366728923000767

Received: 22 February 2023
Revised: 10 September 2023
Accepted: 11 October 2023
First published online: 10 January 2024

Keywords:
bilingualism; questionnaires; content-overlap; Jaccard; consistency

Corresponding author:
John A. E. Anderson;
Email: johnanderson3@cunet.carleton.ca

Abstract

Bilingualism is a multifaceted experience that researchers have examined using various questionnaires to gain insights and characterize the experience. However, there are several issues related to questionnaire choice. To address this, we applied Content Overlap Analysis to seven prevalent bilingualism questionnaires, assessing their affinity. We found little overlap in these questionnaires; most had fewer than 15% of items in common, suggesting they capture different aspects of the bilingual experience and provide complementary rather than redundant data for researchers. Our investigation highlights the importance of choosing a bilingualism assessment tool to carefully fit research questions and sample language experiences.

Introduction

When researchers say they study "bilingualism," there is an assumption that they study the same phenomenon. Yet, bilingualism encompasses various life experiences that are complex, interacting, and dynamic across the lifespan. To document and describe this experience, researchers design questionnaires and definitions focusing on different aspects of bilingual experiences (e.g., age of acquisition, usage, proficiency, dominance, preference). The piecemeal approach is not concordant with how "bilingualism" is broadly used to describe the ability to use multiple languages (Grosjean & Li, 2013). In this paper, we assess the tools that are prevalently used to document adult bilingual experiences. Importantly, we apply a new method to evaluate the magnitude of consensus across these tools. The purpose is to highlight the strengths and diversity of the tools for researchers to evaluate and choose the appropriate tool for their research.

Debates around measuring bilingualism have been ongoing since the field's inception. For example, Grosjean (1989) critiqued the prevailing ratio-based method employed by neurolinguists and psychologists and argued that there are contextual differences in how and when multiple languages are used (also see Complementarity Principles described in Grosjean, 2016). Similarly, Green and Abutalebi (2013) pointed out multiple language contexts for linguistically diverse individuals – suggesting many different sides to bi/multilingual identities beyond pure linguistic knowledge ratios. Building on these ideas, Titone and Tiv (2022) have recently suggested that Bronfenbrenner's (1992) ecological systems approach, which examines the bilingual experience from microscale to large-scale temporal levels, should be implemented in bilingualism research. Though some current questionnaires collect relevant information on a person's bilingual experience or context, questionnaires designed for different studies or locales may highlight diverse experiences.

There has been renewed interest in defining bilingualism more precisely and reaching a consilience among bilingualism questionnaires and researchers. Kašćelan et al. (2022) identified differences in operationalization, components, scale types, and precision in a recent content analysis of bilingualism questionnaires in children. A Delphi census surveying researchers and practitioners also noted the inconsistency in this area (De Cat et al., 2023). While it is logical to propose a new questionnaire that is as comprehensive as possible, creating another ultimate bilingualism questionnaire is perhaps not as helpful as identifying shared and unique aspects of *existing* questionnaires, so researchers can make informed choices about which instrument to use for their research questions (see, for example, this editorial to a special issue in Luk & Esposito, 2020).

To identify the commonalities and uniqueness of existing questionnaires, we apply a quantitative content overlap analysis of existing bilingualism questionnaires. This approach was first used by Fried (2017), who examined widely used measures of depression and concluded that

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

most varied considerably in how they measure the same underlying constructs. Our goal was to identify common items across bilingualism questionnaires to assess their coverage and breadth as indications of consensus and diversity. In our overlap analysis, we identified 50 unique categories of bilingualism items in the seven questionnaires. We calculated Jaccard indices for each questionnaire pair to quantify the similarity between two sets of items to calculate an overlap score for each bilingualism questionnaire relative to all others (Jaccard, 1912). Fried (2017) states that questionnaires are similar if they have overlapping items. If there is little overlap across questionnaires, then it is likely that they are not quantifying constructs similarly. He notes that this is problematic since it decreases the potential for cross-comparisons and generalizations across populations, such as meta-analyses.

Methods

Search strategy

Similar to Fried (2017) and Kašćelan et al. (2022), studies were identified through a formal search facilitated by *Covidence (Veritas Health Innovation, 2022)*. The search strategy was adapted from Kašćelan et al. (2022); however, terms regarding pediatric populations were excluded. The search was performed in PsycINFO, ERIC, Web of Science, and Scopus. Keywords were combined through the Boolean logic term "AND." Searches were limited to items published by May 3rd, 2022. All references were imported to Covidence. The search yielded 8866 papers, with seven papers included for final extraction (see Figure 1 for full Prisma diagram). The full dataset and keywords can be found on the OSF: https://osf.io/s4qug/?view_only=9e6efe3eca434446a503b1d7ef32cf20.

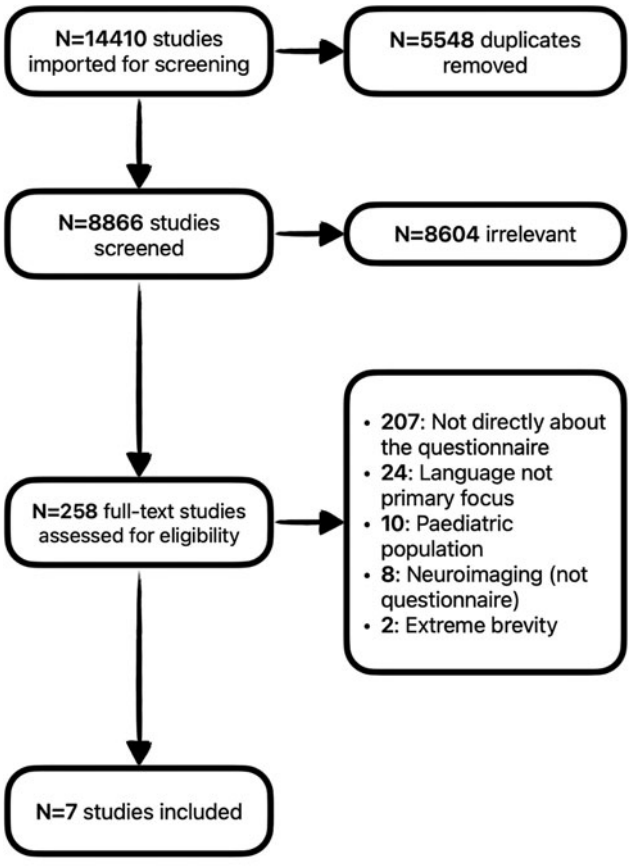


Figure 1. Prisma Diagram.

Eligibility criteria were adapted from Kašćelan et al. (2022). Questionnaires were included if they were in English and mentioned bilingualism. Questionnaires were excluded if they: (1) were parents' reports of children; (2) concerned foreign language learning; (3) concerned speech and language disorders; (4) were about bilingual education; (5) had six or fewer questions about bilingualism; (6) not focused on language; (7) duplicates; and (8) if the full questionnaire could not be obtained. We identified seven questionnaires that are commonly used to capture and quantify bilingualism in adults: the Bilingual Dominance Scale (BDS; Dunn & Fox Tree, 2009), the Bilingualism and Emotion Questionnaire (BEQ; Dewaele & Pavlenko, n.d.), the Bilingual Language Profile (BLP; Birdsong et al., 2012; Gertken et al., 2014), the Bilingualism Switching Questionnaire (BSWQ; Rodriguez-Fornells et al., 2011), the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007), the Language History Questionnaire (LHQ3: Li et al., 2020), and the Language and Social Background Questionnaire (LSBQ; Anderson et al., 2018; Luk & Bialystok, 2013)).

Content analysis

Following Fried (2017), we performed a content analysis on all seven questionnaires. Questions from each questionnaire were prepared in three spreadsheets for different procedures: 1) categorization, 2) within-scale content analysis, and 3) across-scale content analysis.

For categorization, all questions from each questionnaire were placed into one spreadsheet and were organized by category and subcategory (see OSF link above for all categories). For each questionnaire, if a single question contained multiple sub-questions, the sub-questions were treated as separate items and assigned to the appropriate category and subcategory. A total of 222 items across the seven questionnaires were identified and organized into five global categories (*production, switching, exposure, subjective statements, identity, history/acquisition*). These category labels were assigned based on the current understanding of bilingualism components in the existing literature and the aim of the questionnaires (see Table 1 for a description of the questionnaires).

Production referred to questions about how participants used language, including time frequency and the reason for which they used the language (e.g., "If you have children, what language do you speak to them in?; BEQ" or "Estimate how many hours per week you speak language one;" LHQ). Next, switching included questions that explored participants switching or mixing multiple languages (e.g., do you switch between languages with your friends?; "LHQ). Subjective statements included questions that referred to individuals' unique perceptions of their language use (e.g., "how comfortable are you speaking language one?;" LHQ, "what language do you feel dominant in?"; BDS). The identity category is distinct from subjective statements as it only includes questions related to the relationship between one's identity and language use (e.g., "I feel like myself when I speak language one;" BLP or "Which culture/language do you identify with more?"; LHQ), rather than their general feelings about language use. Further, exposure and history/acquisition are similar categories. We elected to categorize questions that were centered around duration and frequency of language exposure as "exposure" – for

https://doi.org/10.1017/S1366728923000767 Published online by Cambridge University Press

Table 1. Questionnaire descriptions

Questionnaire	Purpose	Structure	Length	The practical difficulty of administering
Language History Questionnaire (LHQ 3.0)	To obtain an overview of one's language acquisition and current use.	Fixed scheme (Multiple choice + Likert scale)	27 distinct questions	Comprehensive, however, time consuming
Language Experience and Proficiency Questionnaire (LEAP-Q)	To categorize different bilingualism proficiencies	Fixed scheme (Multiple choice + Likert scale) & open ended	6 distinct questions	Good practicality
BIlingualism Switching Questionnaire	To assess language dominance	Fixed scheme (Multiple choice + Likert scale)	7 distinct questions	Good practicality
BIlingualism Emotion Questionnaire	To understand the relationship between language and emotion	Fixed scheme (Multiple choice + Likert scale)	25 distinct questions	Comprehensive, however, time consuming
Language and Social Background Questionnaire	To understand bilingualism within a socio-cultural context	Fixed scheme (Multiple choice + Likert scale)	22 distinct questions	Comprehensive, however, time consuming
BIlingual Language Profile	To assess language dominance	Fixed scheme (Multiple choice + Likert scale)	19 distinct questions	Good practicality
Bilingual Dominance Scale	To assess language dominance	Fixed scheme (Multiple choice + Likert scale)	8 distinct questions	

example, "Please list the percentage of time you are currently exposed to each language?" (LHQ) – whereas questions focusing on historical language acquisition such as "at what age did you learn the following languages?" (BEQ) were categorized as "history/acquisition." Therefore the difference is about the kind of information that would be elicited – in the first case, the question would focus on the duration or amount of language exposure, while in the second case, the question would have pointed to a single date of language acquisition. Where there were overlaps in categories across questionnaires, these were resolved through consensus among all authors.

After initial categorization was complete, a within-questionnaires overlap analysis was conducted to determine whether questionnaires contained multiple questions assessing the same constructs. Following Fried (2017), if questions were worded similarly or in reverse, they were categorized as one item. This reduced the Bilingual Language Profile (BLP) from 19 items to 12, the Language Experience and Proficiency Questionnaire (LEAP-Q) from 6 to 5, and the BDS from 11 to 8. Overall, this reduced the total number of items from 222 to 211. Next, we conducted the across-scale content analysis. Taking the established list of categories and subcategories, we organized all questions from each questionnaire into one master spreadsheet to identify how frequently each category appeared in each questionnaire. Using information from the across-scales analysis, two construct tables were used to produce two matrices (A and B). Matrix A had three codes: an item was coded "2" if it was specifically featured in the scale, "1" if it was generally featured in the scale, or "0" if it was not featured in the scale. The second matrix, Matrix B, had two codes: an item was coded "1" if it was featured in the scale or "0" if it was not. Twenty-seven idiosyncratic categories were identified.

Statistical analyses

Following Fried (2017), Jaccard similarity indices were calculated in R (R Core Team, 2022), adapting the code Fried shared (see Fried, 2017 for his OSF link). For each pair of questionnaires, the Jaccard index was calculated by computing the ratio of the number of shared items to unique items and shared items. The formula, as described by Fried, is shared/(unique1 + unique2 + shared), where "shared" refers to items shared between the two questionnaires, and unique1 and unique2 refer to the items that are unique to each questionnaire, respectively. We follow Fried's (2017) interpretation of Jaccard index strength which he adapted from Evans (1996) for correlation coefficients very weak 0.00–0.19, weak 0.20–0.39, moderate 0.40–0.59, strong 0.60–0.79, and very strong 0.80–1.0.

Vignettes

All seven questionnaires were coded in Qualtrics and administered to lab members in JAEA's research group. The order of the questionnaire presentation was randomized. Vignettes were first analyzed within each questionnaire. As scoring manuals for questionnaires were unavailable, the authors devised a standardized scoring sheet across all questions (see OSF link for full explanation). All scores were set to fit within the Likert scale values in a given questionnaire. Additionally, some items within a questionnaire were not accounted for if they provided context rather than objectively measuring bilingualism. All scores were re-scaled from original values to 0 (least bilingual) to 1 (most bilingual) to aid visual comparison. Vignette analyses were included to illustrate the similarities and differences in categorization across distinct categories. Given the small sample size (see results for a description of participant demographics), this analysis was meant to serve solely as an example of our findings from the content analysis rather than a comprehensive sample of the bilingual population.

Results

Content overlap

Two hundred and eleven items from seven bilingualism questionnaires were organized into 51 categories. On average, categories appeared 1.607 times in the seven questionnaires (mode=1, median=1). None of the established categories appeared in all questionnaires. The most common categories were production (math) and subjective statements (speaking), which both

https://doi.org/10.1017/S1366728923000767 Published online by Cambridge University Press

Table 2. Category breakdown by Questionnaire

Scale	Categories captured (No.)	Adjusted questionnaire length (No.)	Idiosyncratic categories (%)	Specific categories (%)	Compound categories (%)
LEAP-Q	2	5	0	4	0
BLP	7	12	2	16	0
BEQ	10	55	4	33	0
LHQ	27	8	29	53	4
BSQ	8	13	2	4	0
BDS	9	8	4	16	0
LSBQ	17	28	12	53	2

appeared in four questionnaires (production (math): LHQ, BEQ, BDS, and BLP; subjective statements (speaking): LHQ, BEQ, BDS, LEAPQ). The LHQ was present across the largest number of categories (27), having the most coverage. See Table 2 and Figure 2A for representations of how many times each questionnaire captured a certain category.

Table 2 and Figure 2A summarize the amount of idiosyncratic, compounded, and specific categories, as well as the total adjusted number of items per questionnaire. The LHQ has the largest number of idiosyncratic questions, mainly because it contributes to the MOST categories (27) and contains the most items (88), meaning that it covers constructs that other questionnaires do not. The LSBQ has the second largest number of idiosyncratic categories (12); however, it is also a longer questionnaire (28 items and 17 categories). The remainder of the questionnaires fall within the range of 2-4 idiosyncratic categories, except for the LEAP-Q, which contained no idiosyncratic categories; however, it is also a shorter questionnaire (5 items), meaning it covers fewer categories (2).

The Jaccard Index was next used to estimate the overlap between questionnaires. The average overlap was 0.09, which indicates minimal content overlap across questionnaires (individual and mean overlap across questionnaires are presented in Table 3). Notable elements from the overlap table include:

1. The LENGTH of questionnaires affected the amount of overlap across questionnaires. For example, the LEAP-Q has the fewest number of items (5) and a low overlap (0.04). This could indicate that questionnaires with too few items may need more categories to capture all concepts in the field adequately.
2. The PURPOSE of a questionnaire can affect its overlap with other questionnaires. For instance, the BSQ was explicitly created to look at switching in bilingualism and has a mean overlap of 0.03.
3. Interestingly, the BEQ was also created for a specific purpose (to investigate emotion in bilingualism), yet it has the highest mean overlap of 0.15. It has many items (55); however, it does not appear to have the same low overlap rates as the LSBQ (28 items, 0.10 overlap) or the LHQ (88 items, 0.10 overlap).
4. The highest Jaccard index for any two questionnaires was shared between the BEQ and BDS, which had an overlap score of 0.25, which still constitutes weak overlap.
5. The LEAP-Q and BSQ had the lowest individual overlap with other questionnaires.

To get a sense of how much questionnaires agreed, at least on conceptual categories, we eliminated subcategories and recomputed the Jaccard Index, which rose to 0.42. However, such a broad approach is different from how this measure has traditionally been applied and does not capture item-wise agreement of the questionnaires. Thus, our initial results are the most reliable.

Vignette overview

Six participants (M= 2, F= 4), aged 22-64, completed the questionnaires. Participants were primarily students, with education levels ranging from bachelor's to doctoral degrees and parental education levels from high school to doctoral. Participants mainly resided in the Ottawa region, with some having immigrated to Canada from 1989-2009. The sample was diverse, with ethnicities including White, South Asian, Middle Eastern, and Indo-Caribbean.

We first present each participant's performance within each questionnaire. The purple line represents the participant categorized as "most bilingual," and the green line represents the "least bilingual" individual among the six. As shown in Figure 3, "the most" and "least" bilingual participant differs between questionnaires. The most bilingual participant in the LHQ and BEQ is participant 6, and in the BDS, it is participant 4. Conversely, the most bilingual person on the LEAP-Q, LSBQ, BSQ, and BLP is participant 1. The least bilingual participant in the LHQ, LSBQ, BSQ, and BEQ is participant 3; for the BDS and LEAP-Q, this is participant 6; and for the BLP, it is participant 5. Interestingly, participant 6 was categorized as the *least* (BDS) and *most* (BLP) bilingual participant in different questionnaires. It should be acknowledged that 3 out of 6 participants scored as the least bilingual on BDS, very likely due to the binary nature of scoring items in that questionnaire.

Discussion

Given the diversity of tools characterizing bilingual experiences in adults, we identified the commonality and uniqueness across seven prevalently used language questionnaires. The most prominent finding is that the mean overlap amongst all questionnaires was quite low (0.09). However, differences in length and purpose of questionnaires may have accounted for this effect. This low overlap is evidence that researchers in the field of bilingualism should not use questionnaires interchangeably and must carefully select the most appropriate questionnaire for their selected research questions and the multilingual experiences of their samples. Additionally, the low overlap suggests that it may be inappropriate to use meta-analysis to synthesize data across

https://doi.org/10.1017/S1366728923000767 Published online by Cambridge University Press

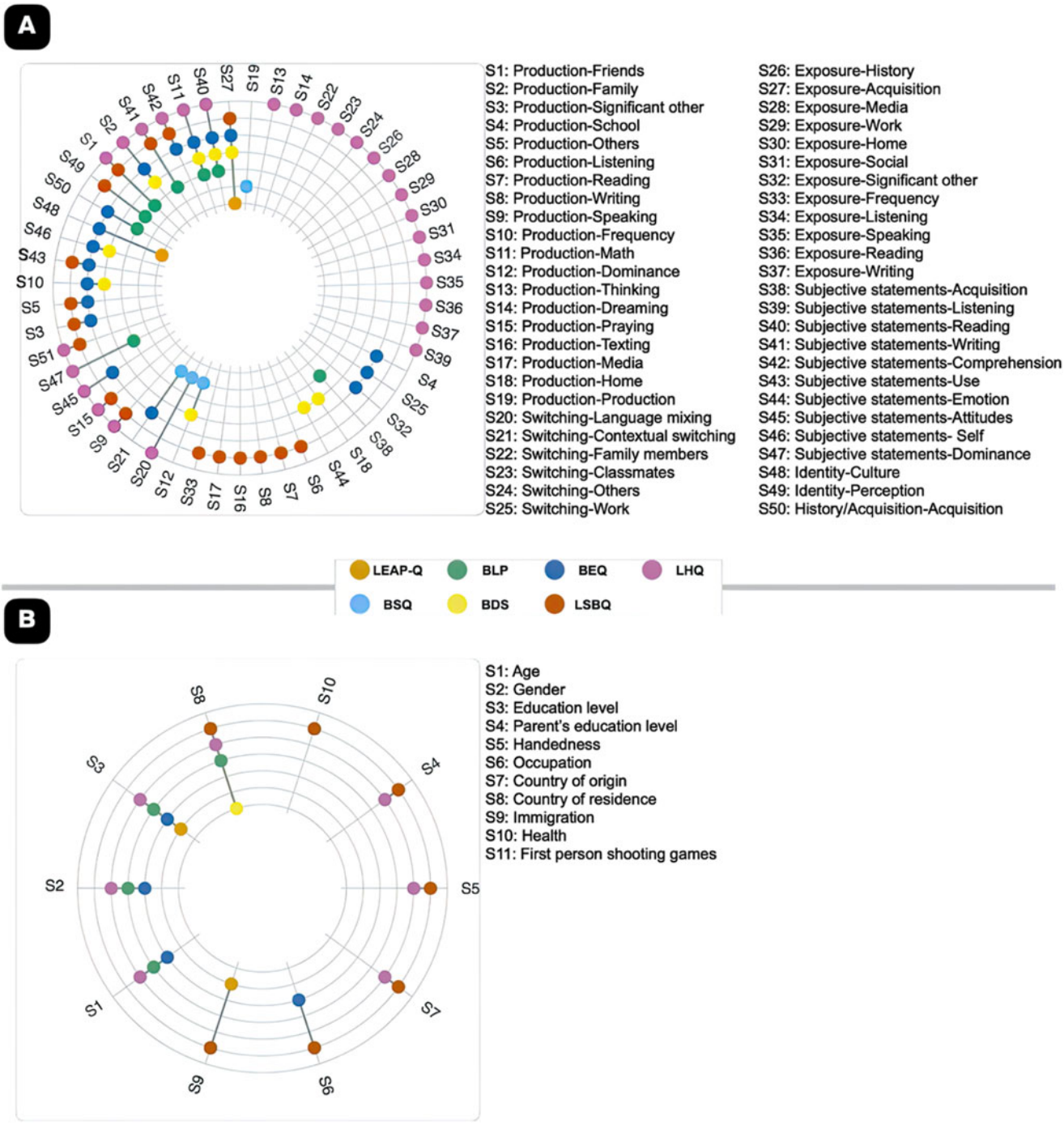


Figure 2. Content Overlap figures for A) language items and B) demographic items from language background questionnaires. LEAP-Q = Language Experience and Proficiency Questionnaire, BLP = Bilingual Language Profile, BEQ = Bilingualism and Emotion Questionnaire.

multiple studies that have characterized bilinguals using different questionnaires. The lack of overlap in questionnaires was also demonstrated in the vignette analysis, as participants were classified differently in individual questionnaires.

Titone and Tiv (2022) proposed a Bilingualism Systems Framework that encompasses three layers of bilingualism that account for interactions, interpersonal, and social dynamics. However, questions that elicit information about these layers and contexts are rarely included in bilingualism questionnaires. Indeed, the investigation of context is often a secondary consideration in these types of scales (Surrain & Luk, 2017). Researchers currently have to try and glean this information from questions designed to measure something else (e.g., construct: switching, questionnaire: LHQ, item: "if you used mixed language in daily life, please indicate the languages that you mix and estimate the frequency of mixing in normal conversation with the following groups of people…").

Overall, the BEQ appears better poised to address more aspects of Titone and Tiv's proposed framework than other questionnaires. Systems one and two are directly addressed in multiple

https://doi.org/10.1017/S1366728923000767 Published online by Cambridge University Press

Table 3. Jaccard Index of seven bilingualism questionnaires

	BSQ	LHQ	BEQ	BLP	LEAP-Q	BDS	LSBQ
BSQ	1.00	0.03	0.05	0	0	0.09	0
LHQ	0.03	1.00	0.13	0.17	0	0.06	0.18
BEQ	0.05	0.13	1.00	0.14	0.12	0.25	0.21
BLP	0	0.17	0.14	1.00	0	0.14	0.13
LEAP-Q	0	0	0.12	0	1.00	0.11	0.05
BDS	0.09	0.06	0.25	0.14	0.11	1.00	0.04
LSBQ	0	0.18	0.21	0.13	0.04	0.05	1.00
Mean overlap	0.03	0.10	0.15	0.10	0.05	0.12	0.10

items, specifically asking about emotional language use for different types of social relationships, including parent-child, friends, and colleagues. However, layers three (societal values, beliefs, and policies) and four (change over time) do not appear to be directly addressed.

Our vignette analysis revealed that using different questionnaires can yield different conclusions about how bilingual an individual is. Which participant was "most bilingual" or "least bilingual" was inconsistent across questionnaires. For example, in one instance, an individual was classified as the most and the least bilingual participant in different questionnaires (BLP and BDS).

Given the findings, we provide four suggestions when choosing a questionnaire to characterize bilingual experiences in adults. First, the BSWQ is the most appropriate when language switching is the research focus. Second, BEQ provides the most information about the interaction between an individual and the social environment where bilingualism occurs. Third, the BEQ and the LSBQ provide the most coverage of categories and overlap to characterize bilingualism more holistically. Finally, for studies aiming to

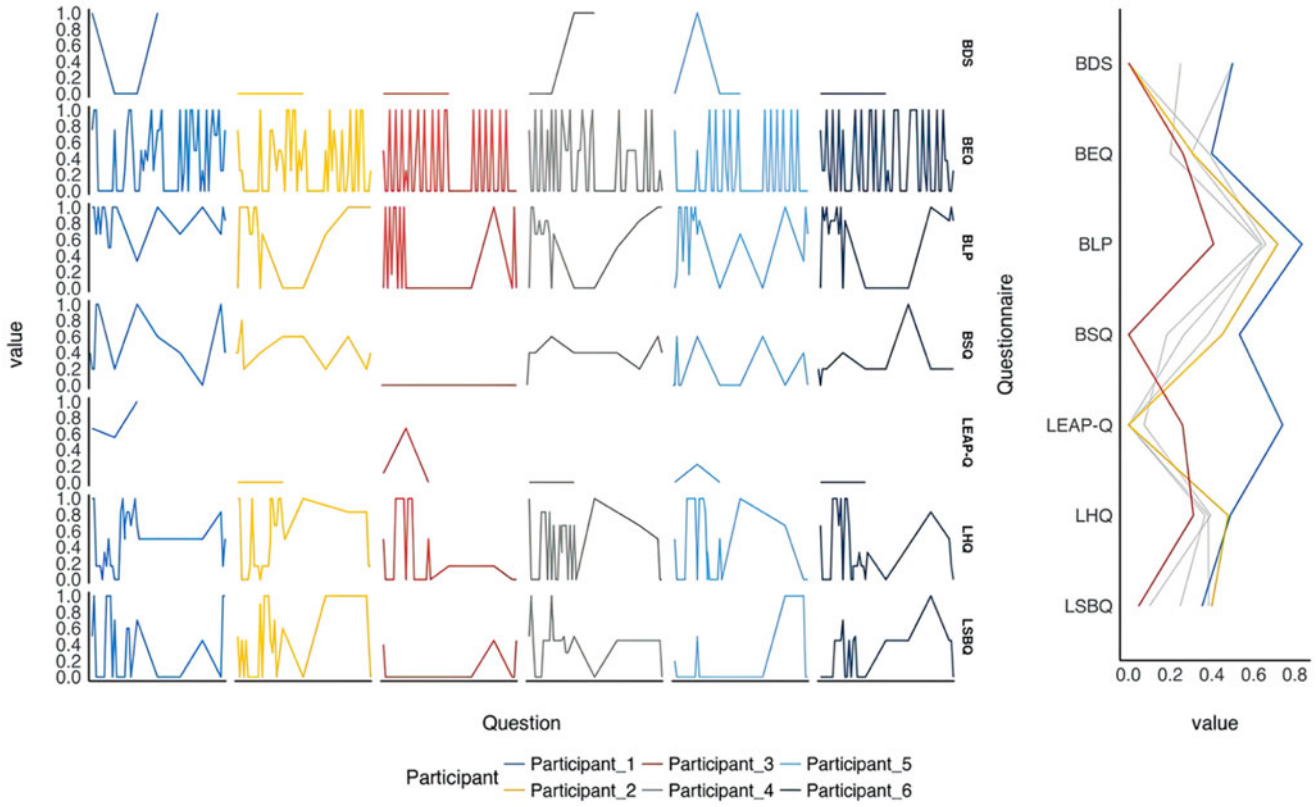


Figure 3. Vignette responses to each of the seven language questionnaires. BDS = Bilingual Dominance Scale, BEQ = Bilingualism and Emotion Questionnaire, BLP = Bilingual Language Profile, BSQ = Bilingualism Switching Quotient, LEAP-Q = Language Experience and Proficiency Questionnaire, LHQ = Language History Questionnaire, LSBQ = Language and Social Background Questionnaire. All numeric items from the above questionnaires were rescaled so that "0" represented monolingualism, and "1" represented higher bilingualism. The x-axis indicates the number of items from each scale, and the lines indicate individual responses. The plots are split so individual item responses can be seen on the left-hand plot, and the average across questions within participants is shown on the right-hand plot. Three of the responses are highlighted in blue, red, and yellow to showcase individuals who are more bilingual (blue) and monolingual (red), as well as an individual who obtained the highest and lowest scores on different measures (yellow).

https://doi.org/10.1017/S1366728923000767 Published online by Cambridge University Press

categorize participants into groups, the BLP provides the most efficient information. Two of the most popular questionnaires are the LEAP-Q and the LHQ. The LHQ3 (Li et al., 2020) offers wide coverage compared to most other questionnaires since its mandate is to incorporate most existing concepts and questions. This breadth may reduce its similarity with more restricted measures, but it remains the most comprehensive questionnaire on our list. The LEAP-Q offers the benefit of being a shorter questionnaire. However, this brevity also limits the LEAP-Q's similarity to the other questionnaires in our study.

The findings of this study must be considered in light of its limitations. A fair critique of these findings is that the amount of agreement will naturally decrease as a function of increasing categories. The parcellation we arrived at is only one perspective on the data. Thus, one way to force more agreement between questionnaires would be to lower the number of categories to the concept level. These categories might, for example, be derived from the Delphi Consensus, which includes: "language exposure and use, language difficulties, proficiency […], education and literacy, input quality, language mixing practices, and attitudes (towards languages and language mixing)" (De Cat et al., 2023). Re-categorizing and analyzing the data from the questionnaires can easily be done using the data we have provided in the OSF. We encourage researchers to explore this dataset from various angles.

The findings of this study highlight some of the issues researchers may have when trying to operationalize bilingualism. Bilingualism is a complex construct, and the existence of multiple tools for capturing various aspects of the experience is not necessarily evidence of a lack of understanding or consensus but rather evidence of the diversity of experience that requires multiple tools to assess properly. Research teams should examine and administer multiple scales to better understand the variations in emphasis and intended use. As such, we caution researchers to carefully choose and supplement tools best suited for their research questions.

Data availability. The data that support the findings of this study are openly available via the Open Science Foundation (OSF) and can be found at https://osf.io/s4qug/?view_only=9e6efe3eca434446a503b1d7ef32cf20.

Funding statement. This research was funded by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), and a Canada Research Chair (CRC, Tier II) to JAEA. R.D. was supported by an NSERC Undergraduate Summer Research Award (USRA).

References

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism. COERLL, University of Texas at Austin. Web. 20 Jan. 2012.*

Bronfenbrenner, U. (1992). Ecological systems theory. In R. Vasta (Ed.), *Six theories of child development: Revised formulations and current issues, (pp* (Vol. 285, pp. 187–249). Jessica Kingsley Publishers. https://psycnet.apa.org/fulltext/1992-98662-005.pdf

De Cat, C., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & The Q-BEx Consortium. (2023). How to quantify bilingual experience? Findings from a Delphi consensus survey. *Bilingualism: Language and Cognition*, *26*(1), 112–124. https://doi.org/10.1017/S1366728922000359

Dewaele, J. M., & Pavlenko, A. (n.d.). Web questionnaire bilingualism and emotions. *University of London*.

Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale*. *Bilingualism: Language and Cognition*, *12*(03)), 273–289. https://doi.org/10.1017/S1366728909990113

Evans, R. H. (1996). An Analysis of Criterion Variable Reliability in Conjoint Analysis. *Perceptual and Motor Skills*, *82*(3), 988–990. https://doi.org/10.2466/pms.1996.82.3.988

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. *Measuring L2 Proficiency: Perspectives from SLA*, *208*, 225. https://books.google.com/books?hl=en&lr=&id=xCLpAwAAQBAJ&oi=fnd&pg=PA208&dq=Bilingual+language+profile+(BLP)&ots=eMKNyf2WVt&sig=eZET4l9PdEugFwzxs0wXbNrkGo8

Green, D. W., & Abutalebi, J. (2013). Language Control in Bilinguals: The Adaptive Control Hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15. https://doi.org/10.1016/0093-934x(89)90048-5

Grosjean, F. (2016). The Complementarity Principle and its impact on processing, acquisition, and dominance. *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, 66–84. https://www.francoisgrosjean.ch/bilin_bicult/12%20Grosjean.pdf

Grosjean, F., & Li, P. (2013). *The Psycholinguistics of Bilingualism*. John Wiley & Sons. https://play.google.com/store/books/details?id=3IumBMf-DJAC

Jaccard, P. (1912). The distribution of the flora in the alpine zone.1. *The New Phytologist*, *11*(2), 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & De Cat, C. (2022). A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition*, *25*(1), 29–41. https://doi.org/10.1017/S1366728921000390

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An Enhanced Tool for Assessing Multilingual Experience. *Bilingualism: Language and Cognition*, *23*(5), 938–944. https://doi.org/10.1017/S1366728918001153

Luk, G., & Bialystok, E. (2013). Bilingualism Is Not a Categorical Variable: Interaction between Language Proficiency and Usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Luk, G., & Esposito, A. G. (2020). BLC Mini-Series: Tools to Document Bilingual Experiences. *Bilingualism: Language and Cognition*, *23*(5), 927–928. https://doi.org/10.1017/S1366728920000632

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research: JSLHR*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rodriguez-Fornells, A., Krämer, U. M., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2011). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, *2*, 388. https://doi.org/10.3389/fpsyg.2011.00388

Surrain, S., & Luk, G. (2017). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, *22*(2), 401–415.

Titone, D. A., & Tiv, M. (2022). Rethinking multilingual experience through a Systems Framework of Bilingualism. *Bilingualism: Language and Cognition*, 1–16. https://doi.org/10.1017/S1366728921001127

Veritas Health Innovation. (2022). *Covidence systematic review software*. www.covidence.org