

cambridge.org/bil

Patrycja Kałamała^{1,2}, Magdalena Senderecka² and Zofia Wodniecka¹

¹Institute of Psychology, Jagiellonian University, Krakow, Poland and ²Institute of Philosophy, Jagiellonian University, Krakow, Poland

Research Article

Cite this article: Kałamała P, Senderecka M, Wodniecka Z (2022). On the multidimensionality of bilingualism and the unique role of language use. *Bilingualism: Language and Cognition* 25, 471–483. <https://doi.org/10.1017/S1366728921001073>

Received: 19 January 2021
Revised: 18 August 2021
Accepted: 8 November 2021
First published online: 10 December 2021

Keywords:

bilingualism; language proficiency; language switching; language entropy; interactional context

Address for correspondence:

Patrycja Kałamała, e-mail – patrycja.kalamala@uj.edu.pl; Zofia Wodniecka, e-mail – zofia.wodniecka@uj.edu.pl; Psychology of Language and Bilingualism Lab, Institute of Psychology, Jagiellonian University in Krakow, ul. Ingardena 6, 30-060 Krakow, Poland

Abstract

The multidimensionality of the bilingual experience makes the investigation of bilingualism fascinating but also challenging. Although the literature distinguishes several aspects of bilingualism, the measurement methods and the relationships between these aspects have not been clearly established. In a group of 171 relatively young Polish–English bilinguals living in their first-language environment, this study investigates the relationships between the multiple measures of bilingualism. The study shows that language entropy – an increasingly popular measure of the diversity of language use – reflects a separate aspect of the bilingual experience from language-switching and language-mixing measures. The findings also indicate that language proficiency is not a uniform aspect of the bilingual experience but a complex construct that requires appropriately comprehensive measurements. Collectively, the findings contribute to the discussion on the best practices for quantifying bilingualism.

1. Introduction

There is a growing understanding in the literature that bilingualism is not a zero-one phenomenon: it is a multifaceted experience, and each aspect should be treated as a continuous rather than binary variable (Antoniou, 2019; Kroll, 2015; Luk, 2015; Luk & Bialystok, 2013; Luk & Esposito, 2020). The fact that the bilingual experience is complex (Grosjean, Grosjean & Li, 2013) makes investigations of bilingualism both fascinating and challenging. In fact, it has been argued recently that it is exactly the complexity of the bilingual experience that is responsible for the “phantom-like quality” of bilingual effects on cognitive functioning (Bialystok, 2017; Blanco-Elorrieta & Pylkkänen, 2018; de Bruin, 2019; Leivada, Westergaard, Duñabeitia & Rothman, 2020; Luk, 2015; Luk & Bialystok, 2013; Luk & Esposito, 2020; Pliatsikas, DeLuca & Voits, 2020). Therefore, efforts have been made in recent years to investigate which aspects of the bilingual experience affect cognitive- and language-related processes. A big challenge in this endeavor is the adequate quantification of the bilingual experience. The goal of this paper is to contribute to the discussion on the best practices for quantifying bilingualism.

Recent reviews (de Bruin, 2019; Surrain & Luk, 2017) indicate that there are three aspects of the bilingual experience that have gained the most prominence in the literature: (1) THE ONSET OF BILINGUALISM, (2) DAILY USE OF LANGUAGES, and (3) LANGUAGE PROFICIENCY (DeLuca, 2019; DeLuca, Segaeert, Mazaheri & Krott, 2020; Leivada et al., 2020). The assessment of the first two aspects is relatively consistent across the research. The onset of bilingualism is typically quantified as the age of second-language (L2) acquisition (L2 AoA; Unsworth, 2016), or the age of active communication in L2 (L2 AoAC; also called the age of active L2 use; Luk, Sa & Bialystok, 2011), although the latter measure is less frequently used. Daily use of a language (also called language exposure) is typically represented as the proportion of time spent using a language. These measures are typically collected using well-established language-background questionnaires such as the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld & Kaushanskaya, 2007; see also Kaushanskaya, Blumenfeld & Marian, 2020), Language History Questionnaire (Li, Zhang, Tsai & Puls, 2014; see also Li, Zhang, Yu & Zhao, 2020), or the Language and Social Background Questionnaire (Anderson, Mak, Keyvani Chahi & Bialystok, 2018). The situation gets more complicated when one wants to quantify language proficiency. While most research has assessed it using self-reports (i.e., asking participants to self-assess how proficient they are in their languages; as implemented, e.g., in the Language Experience and Proficiency Questionnaire, Marian et al., 2007), it has recently been argued that language proficiency is a complex construct and self-assessment may be insufficient to provide a complete indication of bilingual language competence (de Bruin, 2019; DeLuca, Rothman & Pliatsikas, 2019b). This view is supported by a recent study (de Bruin, Carreiras & Duñabeitia, 2017), in which four different measures of language proficiency, i.e., self-assessment, oral proficiency interview, and two language-proficiency tasks (LexTALE, Lemhöfer & Broersma, 2012, and Multilingual Naming Test, Gollan,

© The Author(s), 2021. 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.

CAMBRIDGE
UNIVERSITY PRESS

Weissberger, Runnqvist, Montoya & Cera, 2012), moderately correlated with each other. On the one hand, self-assessment – by sharing variances with the other measures – can be considered a reliable measure of language proficiency. On the other hand, the unique variances of these measures suggest that each reflects a slightly different aspect of language proficiency. Therefore, the literature suggests that self-assessment should be supplemented with additional language proficiency measures. A more comprehensive evaluation of language proficiency is also motivated by recent research which shows that self-assessed language proficiency depends on the personal/cultural characteristics and language-learning history of bilinguals (Hansen, Łuniewska, Simonsen, Haman, Mieszowska, Kofak & Wodniecka, 2019; Tomoschuk, Ferreira & Gollan, 2019). For example, Tomoschuk and colleagues (2019) showed that the correlations between self-assessed language proficiency and an objective language proficiency task (Multilingual Naming Test; Gollan *et al.*, 2012) varied considerably depending on language dominance (see Analysis 3 in the paper). Therefore, self-assessed language proficiency may not be easily generalizable when individuals differ in their bilingual experience.

In addition to the three aspects mentioned above, recent proposals hold that it is also important to consider THE WAYS in which bilinguals use their languages in everyday communication (de Bruin, 2019; DeLuca, 2019; DeLuca, Rothman, Bialystok & Pliatsikas, 2019a; Green & Abutalebi, 2013; Pliatsikas *et al.*, 2020; Wodniecka, Casado, Kałamała, Marecka, Timmer & Wolna, 2020). Two early bilinguals with a native-like proficiency in L2 and similar daily exposure to L2 can still differ in how they actually use their languages. Therefore, THE BILINGUAL PATTERNS OF LANGUAGE USE have been proposed as a fourth important aspect of bilingualism that should be considered. This view is supported by recent theoretical perspectives on bilingual language control (Abutalebi & Green, 2016; Blanco-Elorrieta & Pykkänen, 2018; DeLuca, 2019; Green & Abutalebi, 2013). In particular, the Adaptive Control Hypothesis (Green & Abutalebi, 2013; ACH hereafter) proposes that different patterns of language use may impose completely different demands on bilingual language control and, in turn, differently impact bilinguals' neurocognitive efficiency. Most importantly, the ACH predicts that extensive language switching (i.e., switching that occurs between utterances) in the absence of language mixing (i.e., mixing elements, e.g., words, of two languages within utterances) has the greatest impact on language/cognitive mechanisms in bilinguals.

Although the role of language-use patterns in shaping the bilingual experience has been well-motivated theoretically, their measurement methods are not firmly established in the literature (for reviews, see van den Noort, Struys, Bosch, Jaswetz, Perriard, Yeo, Barisch, Vermeire, Lee & Lim, 2019; Wodniecka *et al.*, 2020). In most studies, language-use patterns were quantified as the frequency of language switching or the frequency of language mixing. A commonly used tool to obtain these measures is the Bilingual Switching Questionnaire (Rodríguez-Fornells, Krämer, Lorenzo-Seva, Festman & Münte, 2012). In addition to these measures, it has also recently been proposed to represent language-use patterns by means of entropy (Gullifer & Titone, 2019; Hartanto & Yang, 2020; Kałamała, Szewczyk, Chuderski, Senderecka & Wodniecka, 2020). The general concept of entropy comes from information theory (Shannon, 1948) and provides a measure of the diversity (uncertainty) of a phenomenon when the relative proportion of occurrences of a set of 'states' is known. When applied to language-use data, these 'states' represent 'languages',

and the entropy value reflects the relative balance (diversity) in the use of different languages. If all languages are used equally, then the entropy value is high; when only a single language is used, then the entropy value is zero.

Summing up, the bilingual experience can be measured in multiple ways. The multitude of available measures is problematic as it is not clear which measures should be chosen to ensure the most adequate quantification of bilingualism. Moreover, the existing variety of measures leads to fragmentation of the literature and makes between-study comparisons difficult as different research groups prefer to use different measures. The most straightforward way to address this issue is to investigate the relationships between the available measures. Such evidence would benefit the literature at both the methodological and the conceptual levels. At the methodological level, it would help to identify those measures that are redundant and those which simply do not reflect inter-individual variability in the bilingual experience. At the conceptual level, it would inform how different measures relate to each other and to what extent they reflect the assumed aspects. However, while many studies have tested how the above-mentioned measures of the bilingual experience relate to (neuro) cognitive functioning (for reviews, see Leivada *et al.*, 2020; van den Noort *et al.*, 2019), surprisingly few studies have tested their interrelationships (Anderson *et al.*, 2018; Fishman & Cooper, 1969; Friesen, Luo, Luk & Bialystok, 2015; Gullifer & Titone, 2019; Gullifer, Kousaie, Gilbert, Grant, Giroud, Coulter, Klein, Baum, Phillips & Titone, 2020; Luk & Bialystok, 2013; Saito, 2015).

Two consecutive studies by Gullifer and colleagues (Gullifer *et al.*, 2020; Gullifer & Titone, 2019) are of particular importance as they took into account not only the classic aspects of the bilingual experience (i.e., onset of bilingualism, daily use of languages, and language proficiency) but also the patterns of language use (assessed by means of language entropy). In both studies, the researchers analyzed data collected via self-assessment from young adult bilinguals living in Montreal. In the second study, the self-assessment of language proficiency was supplemented by a verbal fluency task. In both studies, the language entropy data were collected by asking participants to assess the extent to which they use their languages in different communicative situations, such as when spending time at home or reading. Language entropy estimates were calculated for each type of situation and then subjected to data-reduction techniques (i.e., principal component analysis in the first study and factor analysis in the second one). Data-reduction techniques were used to reduce the complexity of the language entropy data by capturing the common variance of different situational contexts in a smaller set. The first study by Gullifer and Titone (2019) showed that earlier L2 AoA and greater daily use of L2 (called L2 exposure in their paper) were related to better self-assessed L2 abilities. Crucially, the language entropy components (i.e., GENERAL ENTROPY representing language use for personal purposes and WORK ENTROPY representing language use for professional purposes) predicted L2 proficiency over and above the classic measures (i.e., L2 AoA and daily use of L2). Higher language entropy was related to better self-assessed L2 abilities. Most of these effects were then replicated in the second study (Gullifer *et al.*, 2020), in which L2 proficiency was represented as a latent variable comprising fluency tasks in addition to self-ratings. Overall, the picture emerging from Gullifer and colleagues' research is that bilinguals who are highly proficient in L2 acquire L2 early, use L2 frequently, and exhibit great variation in daily language use. While relationships between L2 AoA, daily use of L2, and L2 proficiency have

been reported in several previous studies (Anderson et al., 2018; Luk & Bialystok, 2013), Gullifer and colleagues were the first to show the unique role of patterns of language use in shaping the bilingual experience.

Since the seminal paper by Gullifer and Titone (2019) was published, there have been many calls in the literature for the use of language entropy in characterizing the bilingual experience (de Bruin, 2019; DeLuca, 2019; DeLuca et al., 2019a; DeLuca et al., 2019b; Leivada et al., 2020; Pliatsikas et al., 2020). However, some researchers have recently questioned whether language entropy can actually capture all the complexity of language-use patterns (de Bruin, 2019; Leivada et al., 2020). While the properties of entropy unarguably make it possible to reflect the diversity/balance of language use, it is unclear how well this measure reflects bilingual language-switching habits. Although a high language entropy value seems to imply intense language switching, an equally likely scenario is that a bilingual with a high language entropy score used L2 half of the time but switched between languages only once. Considering that different types of language switching may impose completely different demands on language control (as predicted by the ACH; Abutalebi & Green, 2016; Green & Abutalebi, 2013; Green & Wei, 2014; Green, 2011), it is crucial to learn how the diversity of language use (indexed by language entropy) relates to language-switching habits. So far, however, this relationship has not been studied in the literature.

Moreover, the method of computing the output measures of the language-use patterns proposed in previous research (Gullifer et al., 2020; Gullifer & Titone, 2019) raises a concern. As reviewed above, Gullifer and colleagues proposed estimating language-use patterns on the basis of data-reduction techniques applied to language entropies calculated for different contexts. Data-reduction techniques identify sets of highly correlated variables and extract these as components or factors. When applied to language-entropy data, these techniques decompose unique patterns of language use represented by individual bilinguals into the more general patterns that prevailed in the tested sample. This implies that bilinguals who do not represent typical patterns of language use in a given situation either do not contribute to the factor structure or their contribution is negligible. Consequently, the factorial method does not seem to reflect the unique patterns of bilinguals' language use; instead, it informs about the patterns of language use that prevailed in the tested sample. Moreover, there are two other issues in the research by Gullifer and colleagues (Gullifer et al., 2020; Gullifer & Titone, 2019) that complicate inferences about the role of language-use patterns in shaping the bilingual experience. First, although researchers have collected language-entropy data with regard to many different contexts, some of these contexts pertained to the same spheres of life (e.g., home context and communication with family likely overlap). Since some life spheres were overrepresented in the collected dataset, the observed factor structure could be biased. Second, since not all participants engaged in all targeted contexts, there were missing data in their datasets (three-quarters of participants had missing data in some contexts in the second study). In order to account for the missing data, the researchers artificially imputed NAs. However, this approach is questionable because the lack of data, in this case, is not a technical problem with data collection; instead, it conveys important information that an individual did not engage in a given context. Therefore, artificial data compensation seems to violate the uniqueness of bilinguals' patterns of language use and introduces noise to the results.

1.1. Present study

Although the literature distinguishes several aspects of the bilingual experience, the relationships between these aspects and their measurement methods are not clearly established. The literature urges researchers to further investigate the sources of individual differences and pay particular attention to the role that language-use patterns play in shaping the bilingual experience.

The aim of the presented analyses is to continue recent efforts to quantify the bilingual experience. To this end, we reanalyzed data from our previous study (Kałamała et al., 2020), in which a large sample of relatively young adult bilinguals ($N = 215$ tested in total; $N = 171$ in the analyses) completed a set of questionnaires and two classic L2 proficiency tasks. All participants were Polish-English bilinguals who lived in an L1 environment and declared the daily use of L2. Crucially, the participants displayed great variability in terms of their language experience, but they were relatively homogeneous in terms of their socio-demographic background (for details, see section 2.1). This makes this dataset exceptionally useful as it allows the investigation of individual differences in the bilingual experience while reducing the impact of potential confounding variables related to socio-demographic characteristics. This large dataset allowed us to compute many measures that tapped into the four aspects of the bilingual experience: THE ONSET OF BILINGUALISM (L2 AoA; L2 AoAC), L2 PROFICIENCY (self-ratings for four basic L2 skills; score in the LexTALE task; score in the semantic fluency task), DAILY USE OF L2 (percentage of time spent using L2 each day), and PATTERNS OF LANGUAGE USE (language entropy; language mixing; inter-sentential codeswitching; intra-sentential code-switching).

To examine the relationships between the aspects of bilingualism, we adopted the approach already proposed in the previous literature and asked which aspects of the bilingual experience predict L2 proficiency (Flege, Munro & MacKay, 1995; Gullifer et al., 2020; Gullifer & Titone, 2019). In order to learn about the underlying structure of L2 proficiency, the L2 proficiency data (i.e., self-ratings for listening, reading, speaking, and writing, LexTALE, and semantic fluency scores) were subjected to factor analysis. To preview the findings, the factor analysis showed evidence for two moderately correlated factors. The first factor consisted of self-ratings and was considered to represent participants' subjective beliefs about their L2 abilities and their personal confidence in using this language (SELF-CONFIDENCE IN USING L2 hereafter). The second factor consisted of the LexTALE scores and semantic fluency scores. Since both LexTALE and semantic verbal fluency have been recognized as measuring vocabulary size and lexical access (Gollan, Montoya & Werner, 2002; Lemhöfer & Broersma, 2012; Shao, Janse, Visser & Meyer, 2014), the second factor was considered to represent vocabulary knowledge in L2 (VOCABULARY KNOWLEDGE hereafter). For each latent variable, we fitted a model that included a complementary proficiency measure as well as the other measures of the bilingual experience.

We believe that two main features make the presented analyses unique. First, given the questionable utility of data-reduction techniques in reflecting the uniqueness of language-use patterns, we propose a different approach to estimating indices related to language-use patterns. Here, the patterns of language use were assessed using two questionnaires (one developed for the purpose of the original study and one devised by Hartanto & Yang, 2016), each of which targeted language-use experience with regard to four social contexts (i.e., home, work, school, and free-time

setting). The measures of the language-use patterns (i.e., language entropy, indices of intra-sentential code-switching, index of inter-sentential code-switching) were calculated for each context separately and then averaged across contexts with weights proportional to the time spent using languages in each context. Since the four contexts represented distinct but complementary social settings and we accounted for their relative contribution (using weighted averages), the resulting estimates accommodated the fact that every bilingual represents a unique pattern of language use. Second, the aspects of the bilingual experience were probed using multiple measures, which should inform to what extent the adopted measures reflected the assumed aspects of the bilingual experience.

The overarching goal of this study was to unravel the relationship between language-entropy, language-switching, and language-mixing measures and learn about their utility in differentiating bilinguals in terms of L2 proficiency. This should enable us to propose a more comprehensive description and a more appropriate assessment of the language-use patterns. Moreover, since the performed factor analysis provided evidence for two separable aspects of L2 proficiency – namely, self-confidence in using L2 and vocabulary knowledge – in a post hoc manner we aimed to understand the similarities and differences between these proficiency constructs. In particular, we were interested in identifying those predictor measures that would show the opposite effects for these two aspects of L2 proficiency. Opposite effects for self-confidence in using L2 and vocabulary knowledge in L2 would indicate that predictor measures can simultaneously capture the specific variance of these two proficiency constructs. This, in turn, would indicate the particular utility of these measures in reflecting the bilingual experience.

2. Methods

2.1. Participants

Two hundred and fifteen participants took part in the original study. Only volunteers who declared the daily use of English were invited to the study. Three participants were excluded due to technical problems with data collection. Five other participants were excluded due to the incomplete completion of at least one of the questionnaires. Another 36 participants were excluded because they did not perform the correct language version of either the semantic fluency task or the LexTALE ($n = 22$ and $n = 12$, respectively). In total, one hundred and seventy-one participants were included in the current analyses. The participants' socio-demographic characteristics were assessed using a background questionnaire delivered in Polish (its Polish and English versions can be found at the Open Science Framework, OSF; <https://osf.io/ecnw5/>). On average, the participants were young adults (mean age 24 years, $SD = 4.64$; 132 female; 90 right-handed); they reported a moderate to high income and above-average social status (for details, see Table 1). Their fluid intelligence was measured using a shortened version of Raven's Advanced Progressive Matrices test (only odd-numbered items, 20 minutes to complete; score as the sum of correct responses). On average, participants scored 71% on this test.

Table 2 presents the self-assessment data concerning the participants' language experience (for a description of the questionnaires, see section 2.2). Participants were Polish-English bilinguals who acquired Polish (L1) in their early childhood (below the age of four). Seventeen participants also acquired

Table 1. Participants' socio-demographics based on a self-assessment questionnaire.

Statistic	<i>N</i>	Mean	<i>SD</i>	Min	Max
Age (in years)	171	24	4.64	19	42
Fluid intelligence test score (1–18) ^a	171	12.81	2.63	7	18
Education level (1–4) ^b	171	2.44	0.52	2	4
Parental education level (1–4) ^b	171	2.39	0.68	1	4
Social status (1–10) ^c	171	6.26	1.33	3	9
Income (1–6) ^d	171	2.91	1.18	0	6

Notes. ^ascore in a version of Raven's Advanced Progressive Matrices test (for description see the text); ^bself-ratings were 1 = less than high school, 2 = high school (with/without high school certificate), 3 = college/graduate school (with/without a BS/BS/master certificate), 4 = more than a master's degree; ^cparticipants were asked to self-rate their social status in relation to other people living in their country of residence; they answered on a scale resembling a ladder with rungs numbered from 1 to 10; one rung corresponded to one SES level and only extreme rungs had word labels, i.e., 1 = people who have the least money, the least education, and the least prestigious jobs or no job and 10 = people who have the most money, the most education, and the most prestigious jobs; ^dself-ratings were 1 = less than €500 per month to 6 = more than €4,500 per month; *SD*, standard deviation.

English (L2) in early childhood, while the others started learning L2 in primary school at around the age of seven. On average, the participants started using L2 more intensively when they attended junior high school at around the age of 12. Participants considered their L1 proficiency to be significantly higher than their L2 proficiency, which they considered intermediate to high. They also scored relatively high in English LexTALE (see Table 3). On average, they used L1 for slightly over half the day, and 32% of them used L2 more than L1 on any given day. In addition, 47 participants declared that they had knowledge of some additional languages (predominantly German, Spanish or French). However, the overall proficiency of these additional languages was relatively low, and the daily use of these languages was marginal. They also declared relatively short exposure to L3 (no longer than five years for 34 participants).

The participants lived in their L1 environment (Poland) at the time of testing. The general language environment in Poland is monolingual, with Polish being the official language of communication. However, as globalization continues, the use of other languages (mostly English) is increasing, especially in work environments. This situation was well reflected in the participant sample, as the majority of participants reported using L2 at work (60%). A third of them reported using L2 at school and/or in their free time. Less than 10% of participants reported using L2 at home. L3 experience was limited mainly to school settings (language courses). The participants were likely to use more than one language in the same context (as indicated by relatively high language entropy), and they moderately often switched and mixed languages (based on the indices of language mixing, intra-sentential codeswitching, and inter-sentential codeswitching; for descriptive statistics, see Table 3).

The study met the requirements and gained the approval of the Ethics Committee of Jagiellonian University in Krakow, Institute of Psychology, concerning empirical studies with human participants. The participants were not aware of the reasoning behind the study. Each participant signed an informed consent form prior to the procedure. Following the testing, the participants were debriefed, informed about the study's goals, and paid for their participation (PLN 40, about \$10).

Table 2. Participants' language experience based on a self-assessment questionnaire.

Statistic	Polish (L1) N = 171				English (L2) N = 171				Additional languages (L3-L5) ^a N = 47			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Overall proficiency	8.98	0.12	8	9	7.84	0.91	3	9	5.38	1.75	2	9
Listening	8.99	0.11	8	9	7.87	0.92	3	9	5.75	1.74	2	9
Reading	8.99	0.08	8	9	8.00	0.90	4	9	5.83	1.76	2	9
Speaking	8.98	0.17	7	9	7.76	1.14	2	9	4.98	1.95	1	9
Writing	8.94	0.29	7	9	7.73	1.20	3	9	5.14	2.02	2	9
Age of acquisition^b	0.08	0.41	0	3	6.72	3.41	0	26	14.07	6.22	0	29
Age of active communication^b	1.14	1.92	0	9	11.99	5.16	0	35	17.68	6.27	0	37
% of daily use^c	56	16	1	89	41	15	11	83	10	6	1	25

Notes. SD, standard deviation; the self-ratings range for proficiency were 1 = no knowledge of a given language to 9 = native-like proficiency; ^astatistics for the average of additional languages; ^bage in years; ^cdaily use of languages do not sum up to 100% because not all participants used more than two languages.

Table 3. Correlation matrix (Spearman's rho) and descriptive statistics for the measures of bilingualism.

	use	Lex	SF	listen	read	speak	writ	AoA	AoAC	ent	mix	inter	intra
use	1	0.34	0.07	0.08	0.07	0.12	0.07	0.19	0.12	0.05	-0.04	0.13	0.09
Lex		1	0.47	0.30	0.41	0.28	0.35	0.01	-0.03	-0.13	0.06	0.06	0.13
SF			1	0.29	0.33	0.30	0.34	-0.04	-0.05	0.09	0.06	0.07	0.10
listen				1	0.68	0.67	0.62	-0.15	-0.17	0.18	0.18	0.17	0.21
read					1	0.59	0.68	-0.20	-0.11	0.00	0.15	0.15	0.20
speak						1	0.71	-0.17	-0.19	0.21	0.13	0.16	0.25
writ							1	-0.19	-0.22	0.13	0.23	0.16	0.28
AoA								1	0.60	-0.08	-0.16	-0.21	-0.20
AoAC									1	-0.16	-0.28	-0.27	-0.27
ent										1	0.43	0.42	0.43
mix											1	0.62	0.79
inter												1	0.68
intra													1
Mean	41	79	10.62	7.87	8	7.76	7.73	6.72	11.99	0.81	4.39	4.14	4.83
SD	15	9	5.05	0.92	0.9	1.14	1.20	3.41	5.16	0.28	2.03	1.82	2.11
Min	11	44	2.58	3	4	2	3	0	0	0	0	0	0
Max	83	98	33.68	9	9	9	9	26	35	1.59	8.62	9	9

Notes. use, daily use of L2 in %; Lex, LexTALE score in %; SF, semantic fluency score; listen, listening rating; read, reading rating; speak, speaking rating; writ, writing rating; AoA, age of L2 acquisition; AoAC, age of active communication in L2; ent, index of entropy; mix, index of language mixing; inter, index of inter-sentential codeswitching; intra, index of intra-sentential codeswitching; N = 171; SD, standard deviation; min, minimum; max, maximum; p < .05 bolded.

2.2. Measures

Self-assessment of language proficiency and the onset of bilingualism

Data on the participants' language proficiency and onset of bilingualism were collected using a language background questionnaire, which included questions from two commonly used questionnaires, i.e., the Language Experience and Proficiency Questionnaire (Marian et al., 2007) and the Language History Questionnaire (Li et al., 2014). Participants indicated all languages they know (including their native language and all other

languages learned). Then, they rated their language abilities for each language with regard to listening, reading, speaking, and writing on a scale from 1 (no knowledge of given language) to 9 (native-like proficiency). They also declared the age at which they started acquiring each language and the age at which they started using this language actively in communication. The questionnaire was delivered in Polish, and its Polish and English versions can be found in the OSF repository (<https://osf.io/ecnw5/>). The age of L2 acquisition (L2 AoA) and the age of active communication in L2 (L2 AoAC) served as measures of the onset of

bilingualism. The self-ratings concerning abilities in L2 served as measures of L2 proficiency.

Objective measurement of L2 proficiency

In addition to self-assessment, data on L2 proficiency were collected using two objective proficiency tasks: the Lexical Test for Advanced Learners of English (LexTALE hereafter) and the semantic fluency task.

LexTALE was based on (Lemhöfer & Broersma, 2012). Participants were presented with a series of letter strings, presented one by one on a computer screen in a fixed order. The width of the five-character string on the display was 58 mm (5.0°). For each string, participants indicated whether or not it was an existing English word by pressing either the left shift (yes) or the right shift (no) on the keyboard. If they were unsure, they were instructed to respond “no”. The task was not time-limited. The score was computed as the percentage of correct responses. Before data collection, participants received three practice items to ensure that they understood the task.

The semantic fluency task had a written form. It was based on a version that accommodates individual variation in writing speed (Abrahams, Goldstein, Al-Chalabi, Pickering, Morris, Passingham, Brooks & Leigh, 1997; Abrahams, Leigh, Harvey, Vythelingum, Grise & Goldstein, 2000). Within a 2-min time limit, participants were first asked to produce as many English words as possible that belong to a semantic category (generation condition). Afterward, the list of generated words appeared on the left side of the screen, and participants were asked to rewrite this list on the right side of the screen (control condition). The width of the five-character string on the display was the same as in LexTALE. Participants were instructed to avoid producing repetitions and names of people and places. They gave responses using the keyboard. The control condition was not time-limited, but its duration was measured from the time the first character was written to the time the “STOP” button was pressed. The task included three categories (the same as in (Baus, Costa & Carreiras, 2013; Linck, Kroll & Sunderman, 2009) i.e., fruits and vegetables, animals, parts of the body) that were counterbalanced across participants. The semantic fluency score was calculated as the number of exemplars (without repetitions and proper names) divided by the time spent writing a single word (i.e., the total duration of the control condition divided by the number of words generated in this condition). The main task was preceded by a training session in which participants generated exemplars belonging to the “clothes” category within a 30-s time limit.

Self-assessment of patterns of language use and daily use of L2

Data were collected using two questionnaires: The Patterns of Language Use Questionnaire and The Code-switching and Interactional Contexts Questionnaire. The former questionnaire was developed for the purpose of the original study and was delivered in Polish (for its Polish and English versions, see the OSF repository, <https://osf.io/ecnw5/>). It consists of four parts, each of which targets a different social context: home, work, school, and free time. For each context, participants list all languages they use in this context and estimate how many hours a day they use these languages in that context. If they use more than one language in a given context, they additionally assess how often they mix words of different languages within single utterances. The statements are the same for each context and are accompanied by examples of situations that are specific to a given context. The statements are assessed on a scale from 1

(never) to 9 (always). The theoretical validity of the questionnaire was examined prior to the original study (for details, see Kałamała, Szewczyk, Chuderski, Senderecka & Wodniecka, 2020).

The language-use data served to compute LANGUAGE ENTROPY and THE PERCENTAGE OF DAILY USE OF L2. Language entropy was computed for each context separately using the following formula:

$$\sum_{i=1}^n -p_i \log_2 p_i$$

where p_i is the probability of the use of a language in a context. This procedure resulted in four entropy scores for each participant. Entropies were then averaged across the contexts with weights proportional to the time spent using the languages in each context. A higher score indicates a more diverse use of languages during a given day. If a bilingual uses only one language in a context, then the entropy value is 0. If a bilingual uses several languages to the same degree, then the entropy value is about 1 for two languages and about 1.60 for three languages. Daily use of L2 was computed as the averaged percentage of time spent using L2 during a given day.

The assessment of statements served to compute the INDEX OF LANGUAGE MIXING. Since statement 1 was found to be unreliable in the original study (see Footnote 2 in Kałamała *et al.*, 2020), it was excluded from the computation. The responses to statements 2–4 were first averaged within contexts and then averaged across contexts with weights proportional to the time spent using languages in each context. A lower score indicates less frequent mixing of languages within utterances.

The Code-switching and Interactional Contexts Questionnaire (Hartanto & Yang, 2016) was translated into Polish; the only modification in comparison to the original questionnaire was the use of a 9-point scale (1 = never to 9 = always) instead of a 5-point scale (1 = never to 5 = always) (for its Polish and English versions, see Appendix E in their paper or the OSF repository, <https://osf.io/ecnw5/>). Similarly to the questionnaire by Kałamała and colleagues (2020), the questionnaire consists of four parts, each of which targets one of four social contexts: home, work, school, and free time. For each context, participants assess how often they switch languages between sentences (inter-sentential codeswitching; one question per context) and how often they mix words of different languages within single sentences (intra-sentential codeswitching; one question per context). The INDEX OF INTER-SENTENTIAL CODESWITCHING and THE INDEX OF INTRA-SENTENTIAL CODESWITCHING were computed by averaging corresponding responses with weights proportional to the time spent using languages in each context. For both indices, a lower score indicates less frequent switching between and less frequent mixing of languages. The index of intra-sentential codeswitching corresponds to the index of language mixing from the first questionnaire.

2.3. General procedure

The participants were tested in groups of up to eight during a session lasting approximately 2–2.5 hours (including breaks between blocks of tasks and a longer break in the middle of a session). After informed consent was obtained, the participants performed the battery of tasks and questionnaires, which were administered in Polish in a fixed order: antisaccade task, semantic fluency task, LexTALE task, Stroop task, Patterns of Language Use

Questionnaire, go/no-go task, fluid intelligence test, Code-switching, and Interactional Contexts Questionnaire, stop-signal task, socio-demographic background questionnaire, language background questionnaire. The questionnaires were administered using electronic PDF forms, and the tasks were administered using DMDX (Forster & Forster, 2003). The data concerning the antisaccade task, go/no-go task, Stroop task, and stop-signal task were beyond the scope of this study and have been published elsewhere (Kałamała et al., 2020).

2.4. Data preparation and analyses

The data were analyzed using R (R Core Team, 2019) with the following packages: *stats*, *tidyverse* (Wickham, Averick, Bryan, Chang, McGowan, François, Grolemund, Hayes, Henry, Hester, Kuhn, Pedersen, Miller, Bache, Müller, Ooms, Robinson, Seidel, Spinu, Takahashi, Vaughan, Wilke, Woo & Yutani, 2019), and *psych* (Revelle, 2020). The measures were shown to be reliable (as assessed by split-half reliabilities and Cronbach's α in (Kałamała et al., 2020)). There was no missing data. Most measures demonstrated normal distribution; if a measure had a skewed distribution, it fell within an acceptable range ($|skewness| < 1.80$) (Kline, 2011). All measures were centered and scaled in order to ensure a common measurement scale. Since most of the variables were ordinal, Spearman's rank correlation coefficient (ρ) was used in the correlations.

We followed the analysis workflow of Gullifer and colleagues (Gullifer et al., 2020; Gullifer & Titone, 2019). First, we reduced the complexity of the data. Since the onset-of-bilingualism measures (i.e., L2 AoA and L2 AoAC) and the language-switching measures (i.e., language mixing, inter-sentential codeswitching, intra-sentential codeswitching) highly correlated with each other (ρ s > 0.5 ; see Table 3), we used their averaged values in the analysis (i.e., an averaged onset-of-bilingualism score and an averaged language-switching score). In order to determine the latent structure of language proficiency, the self-ratings concerning the L2 abilities (i.e., listening, reading, speaking, and writing), the LexTALE scores, and the semantic fluency scores were subjected to factor analysis. First, we used the Kaiser-Meyer-Okin test to determine whether the measures were suitable for the factor analysis. A Kaiser-Meyer-Okin score $> .60$ indicated the inclusion of a measure. The number of factors was determined via the parallel function, and the factors were allowed to correlate (i.e., we used oblimin rotation) in order to reflect the real structure of the data. Next, we used multiple regression to determine which factors predict L2 proficiency. For each L2 proficiency factor, we fitted three nested models. The BASE MODEL included the complementary L2 proficiency factors, averaged onset-of-bilingualism score, and daily use of L2. The ADDITIVE MODEL additionally included the language entropy and averaged language-switching score. The INTERACTION MODEL comprised all measures included in the additive model and also all two-way interactions with the onset-of-bilingualism score. Model comparisons using χ^2 tests indicated whether the addition of the measures in the additive and interaction models improved the models' fit. The study materials, data, and R scripts are available at <https://osf.io/ecnw5/>.

3. Results

3.1. Descriptive statistics and correlations

Table 3 presents descriptive statistics and Spearman's ρ correlation matrix for the measures of bilingualism. All of the measures

demonstrated great variability, as indicated by the *SDs* and the min-max value ranges. The measures of L2 proficiency (self-ratings for listening, reading, speaking, and writing, LexTALE score, semantic fluency score) positively correlated with each other, but the strength of the correlation varied between pairs. The onset-of-bilingualism measures (L2 AoA and L2 AoAC) showed a relatively strong positive correlation, suggesting that both could be taken as measures of the same construct. The language-switching measures (i.e., language mixing, inter-sentential codeswitching, intra-sentential codeswitching) also demonstrated strong positive correlations, suggesting that they all reflected language-switching behavior.

3.2. Predicting L2 abilities

The Kaiser-Meyer-Okin test indicated that the data on L2 proficiency (i.e., self-ratings for listening, reading, speaking, and writing, LexTALE, and semantic fluency scores) were suitable for factor analysis (*overall score* = 0.81; no measures fell below the 0.60 threshold for sampling accuracy). The factor analysis for L2 proficiency measures demonstrated that a two-factor solution best described the data and explained 63% of the total variance (see Table 4). The self-ratings loaded onto Factor 1, which explained 74% of the variance. The LexTALE and semantic fluency scores loaded onto Factor 2, which explained 26% of the variance. For Factor 1, the factor loadings were similar. For Factor 2, the LexTALE loaded onto the factor more than the semantic fluency did. Factor 1 was assumed to reflect self-confidence in using L2, whereas Factor 2 constituted an objective measurement of vocabulary knowledge (for a justification, see section 1.2). The factor scores positively correlated with each other ($r = .53$, $p < .001$) and were subsequently used as dependent variables in the regression analyses.

Table 5 presents the outcomes of the regression analyses for self-confidence in using L2 (Factor 1). Compared to the basic model, the addition of language entropy and language switching improved the model's fit, $\chi^2(2) = 4.22$, $p < .05$, but the addition of two-way interactions did not further improve the fit, $\chi^2(4) = 0.60$, $p = .66$. The analysis showed that vocabulary knowledge (Factor 2), the onset of bilingualism, and language entropy significantly predicted self-confidence in using L2. The data suggest that higher self-assessment of L2 proficiency is related to greater vocabulary knowledge, earlier onset of bilingualism, and higher diversity of language use.

Table 6 presents the outcomes of the regression analyses for vocabulary knowledge (Factor 2). Here, the addition of language entropy and language-switching measures to the basic model improved the model's fit, $\chi^2(2) = 6.37$, $p < .001$; the addition of interactions further improved the fit, $\chi^2(4) = 3.93$, $p < .001$. The analysis showed that self-confidence in using L2 (Factor 1), daily use of L2, and language entropy significantly predicted vocabulary knowledge. This suggests that greater vocabulary knowledge is related to higher self-confidence in using L2, greater daily use of L2, and less diverse language use. Compared to the basic model, the interactive model further showed two interactive effects, which are presented in Figure 1. The onset of bilingualism significantly interacted with self-confidence in using L2 (Factor 1), indicating that the relationship between self-assessment and vocabulary knowledge is stronger for bilinguals who became bilinguals relatively early. The onset of bilingualism also interacted with language switching, indicating that more frequent language switching is related to greater vocabulary knowledge, but only for bilinguals

Table 4. Factor analysis for the measures of L2 proficiency.

	Self-confidence in using L2 (Factor 1)	Vocabulary knowledge (Factor 2)	Uniqueness
Listening rating	0.87	−0.07	0.29
Reading rating	0.76	0.14	0.31
Speaking rating	0.85	−0.05	0.31
Writing rating	0.81	0.04	0.32
LexTALE score	−0.02	0.81	0.35
Semantic fluency score	0.14	0.51	0.66

who started using L2 relatively late. At the same time, there was no relationship between language switching and vocabulary knowledge for bilinguals who started using L2 relatively early.

4. Discussion

This study aimed to test the discriminatory properties and inter-relationships between the commonly used measures of the bilingual experience. To this end, we reanalyzed the data from our previous study (Kałamała et al., 2020), in which a relatively large group of young adult Polish–English bilinguals completed a set of language questionnaires and two L2 proficiency tasks. In the analyses, we focused on four different aspects of the bilingual experience and probed each of them using multiple measures. The onset of bilingualism was quantified as the age of L2 acquisition (L2 AoA) and age of active communication in L2 (L2 AoAC); daily use of L2 was represented as the percentage of time spent using L2; L2 proficiency was assessed by means of self-ratings (for listening, reading, speaking, and writing) and two objective L2 proficiency tasks (i.e., LexTALE and semantic fluency). The patterns of language use were assessed by means of language entropy and three different indices reflecting language-switching habits: one reflecting language switching between utterances (i.e., the inter-sentential codeswitching index), and two indices reflecting the mixing of elements of two languages within single utterances (so-called language mixing), i.e., the intra-sentential codeswitching index and the language mixing index. Below, we discuss the relationships between the applied measures and their utility in differentiating bilinguals in terms of L2 proficiency.

4.1. Relationships between the measures of the bilingual experience

The strong relationship between L2 AoA and L2 AoAC suggests that they both tap into the same construct, which we defined as the onset of bilingualism. The strong relationships between the language-switching and language-mixing measures suggest that they jointly reflected a general language-switching tendency in the tested sample: if a bilingual frequently switched languages BETWEEN utterances, they were also likely to mix languages WITHIN single utterances (language-switching tendency hereafter). This finding seems to challenge the widely held notion that language-switching behavior is a continuum in which language switching and language mixing are opposite poles (Deuchar, Muysken & Wang, 2007; Green & Wei, 2014; Hofweber, Marinis & Treffers-Daller, 2016). It should be noted, however, that this

study is the first to compare self-assessed switching and mixing habits in a single study, and it might well be that the relationships between the measures would be different in other language environments, e.g., in a code-switching community. Alternatively, it could also be the case that bilinguals are unable to retrospectively separate their switching and mixing experiences and therefore consider them jointly in self-assessments. We will come back to these issues in section 4.5 (Limitations and future directions). At the same time, language entropy moderately correlated with the language-switching and language-mixing measures, suggesting that language entropy reflects a property of the language-use pattern that is to some extent independent of the language-switching tendencies (this is further confirmed by the regression analyses discussed below). Regarding the measures of L2 proficiency, the factor analysis showed evidence for two moderately correlated factors. The first factor consisted of self-ratings and was considered to represent self-confidence in using L2. The second factor consisted of the LexTALE scores and semantic fluency scores and was considered to represent vocabulary knowledge in L2 (for a justification, see section 1.2).

4.2. Predicting L2 abilities

In order to better understand how the measures differentiate bilinguals in terms of the bilingual experience, we performed a set of regression analyses. Self-confidence in using L2 was related to the onset of bilingualism, indicating that the earlier an individual becomes bilingual, the greater their personal confidence in using L2. Vocabulary knowledge was related to the daily use of L2, indicating that the more bilinguals use L2 on a daily basis, the greater their vocabulary in that language. These effects replicate previously reported findings (Anderson et al., 2018; Friesen et al., 2015; Gullifer et al., 2020; Gullifer & Titone, 2019; Luk & Bialystok, 2013).

The model for vocabulary knowledge further showed the interaction between the onset of bilingualism and self-confidence in using L2 (see the left panel in Figure 1). This interaction indicated that the sooner individuals become bilinguals, the more their personal confidence in using L2 translates into their actual knowledge of the vocabulary of that language. This effect has not been reported in the literature so far, and it has important implications for measuring bilingual language proficiency (we will comment on this in section 4.4).

Crucially, the analyses also showed the discriminative utilities of the measures related to patterns of language use. Language entropy explained the unique variance of both L2 proficiency constructs. A higher value of language entropy was related to better self-confidence in using L2 but worse vocabulary knowledge. These findings indicate that the more diverse the daily language use, the greater the self-confidence in using L2 but the poorer the knowledge of the vocabulary in that language. Moreover, the language-switching tendency was found to be a predictor of vocabulary knowledge but only if quantified by the onset of bilingualism (see the right panel in Figure 1). The data suggest that late bilinguals who frequently switch languages show better knowledge of L2 vocabulary. No such relationship was found for individuals who started using L2 early, which suggests that language-switching tendency improves vocabulary knowledge but only in late bilinguals.

4.3. Implications for measuring patterns of language use

The study demonstrates the important discriminative properties of measures related to patterns of language use. Both language

Table 5. Model outputs for the three nested models that predict self-confidence in using L2 (Factor 1).

	Basic		Additive		Interactive	
Predictor	Estimate [95% CI]		Estimate [95% CI]		Estimate [95% CI]	
(Intercept)	0.00	[-0.12, 0.12]	0.00	[-0.11, 0.11]	-0.01	[-0.13, 0.11]
obj L2 prof	0.57	[0.42, 0.71]	0.59	[0.45, 0.74]	0.60	[0.46, 0.75]
onset	-0.28	[-0.41, -0.15]	-0.25	[-0.39, -0.11]	-0.24	[-0.38, -0.09]
use	0.01	[-0.11, 0.14]	0.01	[-0.12, 0.13]	0.01	[-0.12, 0.13]
switch	-		0.00	[-0.16, 0.15]	-0.02	[-0.18, 0.14]
entropy	-		0.17	[0.04, 0.31]	0.19	[0.05, 0.32]
onset : obj L2 prof	-		-		0.10	[-0.10, 0.30]
onset : use	-		-		-0.06	[-0.19, 0.07]
onset : switch	-		-		-0.14	[-0.33, 0.06]
onset : entropy	-		-		0.06	[-0.11, 0.24]
R^2	.353		.385		.394	

Notes. *obj L2 prof*, objective L2 proficiency score; *onset*, averaged onset-of-bilingualism score; *use*, daily use of L2; *switch*, averaged language-switching score; *entropy*, index of language entropy; *CI*, confidence intervals; $N = 171$; $p < .05$ bolded.

Table 6. Model outputs for the three nested models that predict vocabulary knowledge (Factor 2).

	Basic		Additive		Interactive	
Predictor	Estimate [95% CI]		Estimate [95% CI]		Estimate [95% CI]	
(Intercept)	0.00	[-0.11, 0.11]	0.00	[-0.10, 0.10]	-0.01	[-0.12, 0.09]
sub L2 prof	0.47	[0.35, 0.59]	0.48	[0.37, 0.60]	0.58	[0.46, 0.70]
onset	0.03	[-0.10, 0.16]	0.04	[-0.08, 0.17]	-0.00	[-0.14, 0.13]
use	0.18	[0.07, 0.29]	0.16	[0.06, 0.27]	0.17	[0.06, 0.28]
switch	-		0.13	[-0.01, 0.27]	0.10	[-0.04, 0.24]
entropy	-		-0.21	[-0.32, -0.09]	-0.23	[-0.34, -0.11]
onset : sub L2 prof	-		-		-0.20	[-0.31, -0.09]
onset : use	-		-		0.00	[-0.10, 0.11]
onset : switch	-		-		0.18	[0.02, 0.34]
onset : entropy	-		-		-0.02	[-0.17, 0.13]
R^2	.331		.376		.431	

Notes. *sub L2 prof*, subjective L2 proficiency score; *onset*, averaged onset-of-bilingualism score; *use*, daily use of L2; *switch*, averaged language-switching score; *entropy*, index of language entropy; *CI*, confidence intervals; $N = 171$; $p < .05$ bolded.

entropy and language-switching tendency explained the unique variance of L2 proficiency over and above the classic measures of the bilingual experience, such as the onset of bilingualism and daily use of L2. Importantly, language entropy was related to both L2 proficiency factors but in exactly the opposite way. This indicates that language entropy covers variance that is simultaneously specific to both subjective and objective measurements. This finding is novel and speaks to the importance of accounting for language entropy in research practice: by taking language entropy into account, we are able to establish the conditions under which greater self-confidence in using a language is accompanied by more impoverished vocabulary in that language.

The significant effects of language entropy and language-switching tendency indicate the important role of language-use patterns in shaping the bilingual experience. However, the significance of these effects could also be attributed to the fact that language entropy and language-switching measures accounted for unique variance related to experience of additional languages

(L3), whereas the other predictor measures did not. Since only a minority of the participants reported contact with L3, and the L3-related variance was limited (see section 2.1), we deliberately did not make any reference to this source of variance in the analyses. Nevertheless, to unequivocally reject the hypothesis that L3 experience affected the pattern of results, we fitted additional regression models in which the predictor pertaining to daily use of L2 had been replaced with the overall daily use of additional languages (for a similar analysis, see Gullifer & Titone, 2019). The pattern of results largely remained the same.¹

Moreover, it should be noted that the language-entropy effects reported in this study do not fully correspond to the previously

¹For self-confidence in using L2, the additive model was still the best-fitting model ($\chi^2(2) = 4.14, p < .03$). For vocabulary knowledge, the interactive model also remained the best-fitting model ($\chi^2(4) = 4.00, p < .05$). The significance and direction of the effects were the same as in the original models. The magnitude of some of the effects has changed, but only slightly, and thus did not affect the interpretation provided in the paper.

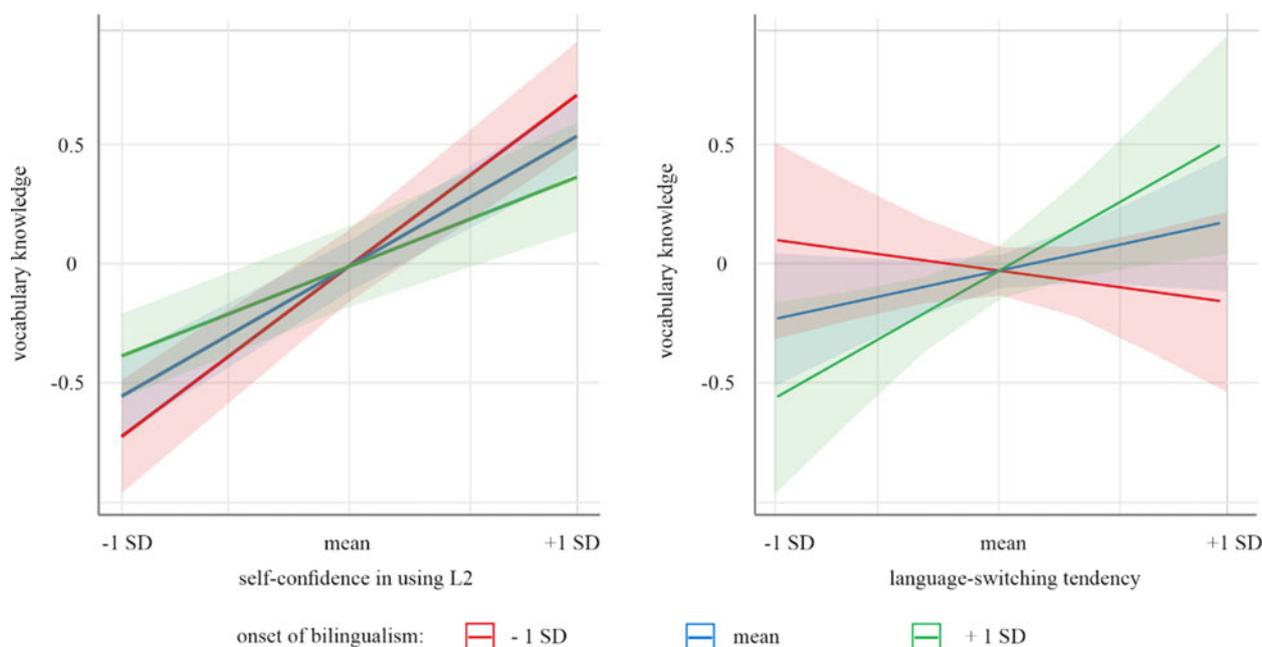


Figure 1. Predicted values of vocabulary knowledge derived from the interactive model. Panel A shows vocabulary knowledge as a function of self-confidence in using L2 quantified by the onset of bilingualism. Panel B shows vocabulary knowledge as a function of the averaged language-switching score quantified by the onset of bilingualism. Vocabulary knowledge and self-confidence in using L2 refer to the respective factor scores. Onset of bilingualism refers to the averaged onset-of-bilingualism score. Language-switching tendency refers to the averaged language-switching score (for descriptions, see the text). *SD* refers to standard deviation. Ribbons represent 95% confidence intervals.

reported findings. Specifically, Gullifer and colleagues reported positive effects of language entropy on language proficiency in two consecutive studies (Gullifer *et al.*, 2020; Gullifer & Titone, 2019). In their earlier study, language proficiency was assessed via self-assessment. In the later study, self-assessment was accompanied by fluency tasks and, thus, was assumed to provide a more objective measurement of language proficiency. Therefore, the negative effect of language entropy for vocabulary knowledge, which in fact was measured using objective proficiency tasks, seems to contradict the second study of this group. We think, however, that the present study cannot be directly compared with previous research as we employed a different approach to estimating overall language-entropy scores. The weighted-average method used in our study considers the overall variability of the language-use patterns represented by individual bilinguals, thus providing a complete and comprehensive quantification. This was not necessarily the case in previous studies because Gullifer and colleagues employed data-reduction techniques that focus on extracting common variance. It may therefore be the case that the specific variance that was accounted for in our analyses was filtered out by the data-reduction techniques used in previous research. Moreover, a closer look at the outcomes of the factor analysis by Gullifer and colleagues suggests that the factor that was assumed to represent an objective measurement of L2 proficiency in their study was dominated by self-ratings, while the L2 fluency tasks made little contribution to this factor. This undermines the objectivity of this measurement and suggests that the discrepancy between the present findings and the findings reported in Gullifer and colleagues (2020) is only apparent: if the measurement of language proficiency mostly reflects self-confidence in using a language, then greater diversity in language use translates into better language proficiency. However, if language proficiency points to vocabulary knowledge, then greater

diversity in language use should translate into poorer language proficiency. This speaks to the importance of the adequate operationalization of language proficiency. We will come back to this issue in the following section.

Summing up, the research by Gullifer and colleagues (Gullifer *et al.*, 2020; Gullifer & Titone, 2019), which showed the predictive value of language entropy, certainly made an important contribution to understanding the sources of individual differences in bilingualism. Our research further advances the state of knowledge, showing that greater diversity of language use is related to better L2 proficiency but only if language proficiency is understood as personal confidence in using L2. The opposite effect should be expected for vocabulary knowledge. Overall, the moderate correlations between language entropy and language switching, along with their unique effects on vocabulary knowledge in the current study, suggests that there might be two interrelated aspects of language-use patterns: one that reflects the balance (diversity) of language use (represented by language entropy), and another that reflects general language-switching tendency (represented by the language-switching and mixing measures). This implies that language entropy should not be considered a substitute for language-switching measures (Gullifer *et al.*, 2020; Gullifer & Titone, 2019); instead, both aspects should be considered to ensure sufficiently precise measurement of language-use patterns.

4.4. Implications for measuring language proficiency

So far, most studies have assessed language proficiency on the basis of self-assessment (Surrain & Luk, 2017). However, as reviewed in the Introduction, self-assessment of L2 proficiency may not provide a complete indication of language proficiency (de Bruin *et al.*, 2017; Tomoschuk *et al.*, 2019). Moreover, limiting the

evaluation of language proficiency to self-assessment can be misleading when bilinguals differ considerably in their language experiences (Tomoschuk et al., 2019). Therefore, the literature calls for a more comprehensive measurement of L2 proficiency. The outcomes of our factor analysis further support this call. Evidence for two moderately correlated factors of L2 proficiency suggests that language proficiency is not a uniform aspect of the bilingual experience but is itself a complex construct. Importantly, self-assessment of language abilities is insufficient to fully account for vocabulary knowledge, and additional measurements are needed to represent this aspect of language proficiency.

The outcomes of the regression analyses further advance the current understanding of the factors that affect language proficiency estimates. Language entropy and the onset of bilingualism seem to play important roles in shaping self-confidence in using L2, whereas language entropy, language switching, and daily use of L2 seem to shape the knowledge of L2 vocabulary (but language switching impacted vocabulary knowledge only in the case of the late onset of bilingualism). Importantly, self-confidence in using L2 was more likely to overlap with vocabulary knowledge for the earlier onset of bilingualism, thus suggesting that the earlier L2 is acquired, the more language proficiency is considered through the prism of vocabulary knowledge. This finding contributes to the discussion on the validity of self-assessed language proficiency. While Tomoschuk and colleagues (2019) pointed to the role of language dominance in interpreting self-assessed L2 proficiency within a bilingual group, we show that bilinguals' understanding of self-assessed language proficiency (i.e., the very meaning behind the concept of language proficiency) may change depending on the onset of bilingualism.

4.5. Limitations and future directions

The study advances the current understanding of the factors that shape the bilingual experience. However, some of the reported findings are limited by the available data and require further investigation. The first limitation is related to the specificity of the tested participant sample. The participants were comparable in terms of their socio-demographic characteristics, which reduced the impact of experiential factors not related to language experience. However, since most of the participants were relatively young adults (88% under thirty years old) and all of them lived in their L1 environment, the findings cannot be generalized to the entire bilingual population. In particular, we anticipate that the relationship between language switching and mixing may be different for older bilinguals or other language environments. Therefore, an important question concerns the extent to which the switching-mixing relationship observed for relatively young bilinguals living in the L1 environment generalizes to other bilingual populations.

The second limitation is related to the selection of measures and the corresponding aspects of the bilingual experience. In this study, we have focused on those aspects that have received the greatest attention in the literature (Marian & Hayakawa, 2021; Surrain & Luk, 2017). However, some researchers have also pointed to other aspects of bilingualism that we have not accounted for in this study (e.g., identity with L2 culture or language environment/context; for more examples, see Marian & Hayakawa, 2021). Therefore, future research should consider addressing a more extensive range of aspects. Moreover, the data indicate that self-assessment of language proficiency may hinder credible comparisons between bilinguals. However, it needs to be acknowledged that eight of our measures were, in

fact, based on self-assessment. Therefore, one may argue that the self-assessment biased the results related to the other aspects of the bilingual experience as well. We think, however, that this is an unlikely scenario. Since the measures related to the onset of bilingualism and daily language use (e.g., L2 AoA, language entropy, daily use of L2) referred to relatively objective facts, they are largely assumed to be reliable and valid (de Bruin, 2019; Leivada et al., 2020; Marian & Hayakawa, 2021). Their reliability was also supported by our additional analysis.² Yet, it should be acknowledged that little is known about the reliability and validity of self-assessed language-switching habits (but see de Bruin, 2019; Dewaele & Wei, 2014; Jevtović, Duñabeitia & de Bruin, 2020). Therefore, the strong correlations between language switching and language mixing should be treated with caution as they could also be attributed to the fact that bilinguals cannot separate their language-switching and language-mixing tendencies and always consider them collectively. Future research should thoroughly examine how well self-ratings reflect actual language-switching behavior across different language environments. This can be achieved by comparing natural language-use habits (retrieved from voice recordings collected over a day) with self-assessment data. Relatedly, we argue that language proficiency is a complex construct and provide evidence for its 'vocabulary knowledge' aspect (for a similar argument, see DeLuca et al. 2019b). Therefore, future studies should investigate whether the effects reported for vocabulary knowledge can be replicated using other proficiency measures, such as the Multilingual Naming Test (Gollan et al., 2012), which is used to measure expressive vocabulary, or the Oxford Quick Placement Test (Geranpayeh, 2003), which is used to measure grammatical knowledge. In order to facilitate future research and enable direct between-study comparisons, we have shared the materials and scripts that we used in this study. The free online-available tool for calculating entropy scores devised by Gullifer and Titone (2018) could also be helpful.

5. Conclusions

The study shows the interrelationships and discriminative utilities of measures related to the bilingual experience in a relatively large group of young adult bilinguals living in their L1 environment. In particular, the study shows the discriminative potential of measures related to language-use patterns, i.e., language entropy and broadly understood language-switching tendency. Overall, the correlation pattern and the outcomes of the regression analyses suggest that language-use patterns can be composed of two inter-related aspects: the balance (diversity) of language use and language-switching tendency. The study also suggests that language switching and language mixing may not always be two sides of the same coin as they can coincide in some bilingual populations (but this effect needs further investigation). Moreover, the study also draws research attention to the problem of measuring language proficiency. The self-assessment of language proficiency is likely to reflect self-confidence in using a language. Crucially, bilinguals' individual understanding of language

²In order to verify the reliability of L2 AoA and L2 AoAC, we performed an additional analysis on a longitudinal dataset collected in our laboratory (63 Polish-English bilinguals; a 7-month period between two subsequent testing sessions). The intra-class correlation coefficients (internal consistency for a two-way fixed model) for the L2 AoA and L2 AoAC were 0.91 and 0.81, respectively, which indicates satisfactory reliability of the measurements (Cicchetti, 2001).

proficiency (the subject of self-assessment) may differ, even within a participant sample drawn from the same language environment. Therefore, self-assessment should not be considered a unique index of language proficiency but should be supplemented with additional measurements.

Acknowledgments. The research was funded by a Ministry of Science and Higher Education grant (0002/DIA/2014/43). During work on the paper, P.K. was supported by National Science Centre grants (2017/27/N/HS6/01029; 2020/36/T/HS6/00363) and a Foundation for Polish Science scholarship (START); Z.W. and P.K. were supported by a National Science Centre grant (2015/18/E/HS6/00428); and M.S. was supported by a National Science Centre grant (2015/19/B/HS6/00341). The authors gratefully acknowledge the help of Jakub Szewczyk with designing the language-entropy measure, Adam Chuderski with consulting the analyses, and Michael Timberlake with proofreading. Upon acceptance, all raw data and the code needed to reproduce all results will be made publicly available.

Competing interests declaration. The authors declare none.

References

- Abrahams S, Goldstein LH, Al-Chalabi A, Pickering A, Morris RG, Passingham RE, Brooks DJ and Leigh PN (1997) Relation between cognitive dysfunction and pseudobulbar palsy in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry* **62**(5), 464–472.
- Abrahams S, Leigh PN, Harvey A, Vythelingum GN, Grise D and Goldstein LH (2000) Verbal fluency and executive dysfunction in amyotrophic lateral sclerosis (ALS). *Neuropsychologia* **38**(6), 734–747.
- Abutalebi J and Green DW (2016) Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition* **19**(4), 689–698. <https://doi.org/10.1017/S1366728916000225>
- Anderson JAE, Mak L, Keyvani Chahia A and 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>
- Antoniu M (2019) The Advantages of Bilingualism Debate. *Annual Review of Linguistics* **5**(1), 395–415. <https://doi.org/10.1146/annurev-linguistics-011718-011820>
- Baus C, Costa A and Carreiras M (2013) On the effects of second language immersion on first language production. *Acta Psychologica* **142**(3), 402–409. <https://doi.org/10.1016/j.actpsy.2013.01.010>
- Bialystok E (2017) The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin* **143**(3), 233–262. <https://doi.org/10.1037/bul0000099>
- Blanco-Elorrieta E and Pylkkänen L (2018) Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences* **22**(12), 1117–1126. <https://doi.org/10.1016/j.tics.2018.10.001>
- de Bruin A (2019) Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences* **9**(3), 33. <https://doi.org/10.3390/bs9030033>
- de Bruin A, Carreiras M and Duñabeitia JA (2017) The BEST dataset of language proficiency. *Frontiers in Psychology* **8**, 7.
- DeLuca V (2019) Future directions in examining neurological adaptation to bilingual experiences. *Journal of Experimental Neuroscience* **13**, 117906951987659. <https://doi.org/10.1177/1179069519876597>
- DeLuca V, Rothman J, Bialystok E and Pliatsikas C (2019a) Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences* **116**(15), 7565–7574. <https://doi.org/10.1073/pnas.1811513116>
- DeLuca V, Rothman J and Pliatsikas C (2019b) Linguistic immersion and structural effects on the bilingual brain: A longitudinal study. *Bilingualism: Language and Cognition* **22**(5), 1160–1175. <https://doi.org/10.1017/S1366728918000883>
- DeLuca V, Segart K, Mazaheri A and Krott A (2020) Understanding bilingual brain function and structure changes? U bet! A unified bilingual experience trajectory model. *Journal of Neurolinguistics* **56**, 100930. <https://doi.org/10.1016/j.jneuroling.2020.100930>
- Deuchar M, Muysken P and Wang S-L (2007) Structured Variation in Codeswitching: Towards an Empirically Based Typology of Bilingual Speech Patterns. *International Journal of Bilingual Education and Bilingualism* **10**(3), 298–340. <https://doi.org/10.2167/beb445.0>
- Dewaele J-M and Wei L (2014) Attitudes towards code-switching among adult mono- and multilingual language users. *Journal of Multilingual and Multicultural Development* **35**(3), 235–251. <https://doi.org/10.1080/01434632.2013.859687>
- Fishman JA and Cooper RL (1969) Alternative measures of bilingualism. *Journal of Verbal Learning & Verbal Behavior* **8**(2), 276–282. [https://doi.org/10.1016/S0022-5371\(69\)80076-9](https://doi.org/10.1016/S0022-5371(69)80076-9)
- Fllege JE, Munro MJ and MacKay IRA (1995) Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* **97**(5), 3125–3134. <https://doi.org/10.1121/1.413041>
- Forster KI and Forster JC (2003) DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* **35**(1), 116–124. <https://doi.org/10.3758/BF03195503>
- Friesen DC, Luo L, Luk G and Bialystok E (2015) Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. *Language, Cognition and Neuroscience* **30**(3), 238–250. <https://doi.org/10.1080/23273798.2014.918630>
- Geranpayeh A (2003) A quick review of the English quick placement test. *Research Notes* 8–10.
- Gollan TH, Montoya RI and Werner GA (2002) Semantic and letter fluency in Spanish–English bilinguals. *Neuropsychology* **16**(4), 562–576. <https://doi.org/10.1037/0894-4105.16.4.562>
- Gollan TH, Weissberger GH, Runnqvist E, Montoya RI and Cera CM (2012) Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition* **15**(3), 594–615. <https://doi.org/10.1017/S1366728911000332>
- Green DW (2011) Language control in different contexts: The behavioral ecology of bilingual speakers. *Frontiers in Psychology* **2**. <https://doi.org/10.3389/fpsyg.2011.00103>
- Green DW and 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>
- Green DW and Wei L (2014) A control process model of code-switching. *Language, Cognition and Neuroscience* **29**(4), 499–511. <https://doi.org/10.1080/23273798.2014.882515>
- Grosjean F, Grosjean F and Li P (2013) Bilingualism: A short introduction. *The Psycholinguistics of Bilingualism* 5–25.
- Gullifer JW, Koussaie S, Gilbert AC, Grant A, Giroud N, Coulter K, Klein D, Baum S, Phillips N and Titone D (2020) Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency. *Applied Psycholinguistics* **42**(2), 245–278. <https://doi.org/10.1017/S0142716420000521>
- Gullifer JW and Titone D (2018) *Compute language entropy with {languageEntropy}*. Retrieved from <https://github.com/jasongullifer/languageEntropy>.
- Gullifer JW and Titone D (2019) Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition* 1–12. <https://doi.org/10.1017/S1366728919000026>
- Hansen P, Łuniewska M, Simonsen HG, Haman E, Mieszowska K, Kołak J and Wodniecka Z (2019) Picture-based vocabulary assessment versus parental questionnaires: A cross-linguistic study of bilingual assessment methods. *International Journal of Bilingualism* **23**(2), 437–456. <https://doi.org/10.1177/1367006917733067>
- Hartanto A and Yang H (2016) Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition* **150**, 10–19. <https://doi.org/10.1016/j.cognition.2016.01.016>
- Hartanto A and Yang H (2020) The role of bilingual interactional contexts in predicting interindividual variability in executive functions: A latent variable

- analysis. *Journal of Experimental Psychology: General* **149**(4), 609–633. <https://doi.org/10.1037/xge0000672>
- Hofweber J, Marinis T and Treffers-Daller J (2016) Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism* **6**(5), 648–668. <https://doi.org/10.1075/lab.15052.hof>
- Jevtović M, Duñabeitia JA and de Bruin A (2020) How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. *Bilingualism: Language and Cognition* **23**(2), 401–413. <https://doi.org/10.1017/S1366728919000191>
- Kalamala P, Szewczyk J, Chuderski A, Senderecka M and Wodniecka Z (2020) Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition* **204**, 104373. <https://doi.org/10.1016/j.cognition.2020.104373>
- Kaushanskaya M, Blumenfeld HK and Marian V (2020) The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition* **23**(5), 945–950. <https://doi.org/10.1017/S1366728919000038>
- Kline RB (2011) *Principles and practice of structural equation modeling* (5th ed.). New York: The Guilford Press, pp. 3–427.
- Kroll JF (2015) On the consequences of bilingualism: We need language and the brain to understand cognition. *Bilingualism: Language and Cognition* **18** (01), 32–34. <https://doi.org/10.1017/S1366728914000637>
- Leivada E, Westergaard M, Duñabeitia JA and Rothman J (2020) On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed? *Bilingualism: Language and Cognition* **1**–14. <https://doi.org/10.1017/S1366728920000358>
- Lemhöfer K and Broersma M (2012) Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* **44**(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Li P, Zhang F, Tsai E and Puls B (2014) Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition* **17**(03), 673–680. <https://doi.org/10.1017/S1366728913000606>
- Li P, Zhang F, Yu A and 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>
- Linck JA, Kroll JF and Sunderman G (2009) Losing access to the native language while immersed in a second language evidence for the role of inhibition in second-language learning. *Psychological Science* **20**(12), 1507–1515. <https://doi.org/10.1111/j.1467-9280.2009.02480.x>
- Luk G (2015) Who are the bilinguals (and monolinguals)? *Bilingualism: Language and Cognition* **18**(01), 35–36. <https://doi.org/10.1017/S1366728914000625>
- Luk G and 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 and Esposito AG (2020) BLC mini-series: Tools to document bilingual experiences. *Bilingualism: Language and Cognition* **23**(5), 927–928. <https://doi.org/10.1017/S1366728920000632>
- Luk G, Sa ED and Bialystok E (2011) Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition* **14**(4), 588–595. <https://doi.org/10.1017/S1366728911000010>
- Marian V, Blumenfeld HK and 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* **50**(4), 940–967.
- Marian V and Hayakawa S (2021) Measuring bilingualism: The quest for a “bilingualism quotient”. *Applied Psycholinguistics* **42**(2), 527–548. <https://doi.org/10.1017/S0142716420000533>
- Pliatsikas C, DeLuca V and Voits T (2020) The many shades of bilingualism: language experiences modulate adaptations in brain structure. *Language Learning* **70**(S2), 133–149. <https://doi.org/10.1111/lang.12386>
- R Core Team (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Revelle W (2020) *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University, Evanston, Illinois. R package version 2.0.12, <https://CRAN.R-project.org/package=psych>.
- Rodriguez-Fornells A, Krämer UM, Lorenzo-Seva U, Festman J and Münte TF (2012) Self-assessment of individual differences in language switching. *Frontiers in Psychology* **2**. <https://doi.org/10.3389/fpsyg.2011.00388>
- Saito K (2015) The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition* **37**(4), 713–743. <https://doi.org/10.1017/S0272263115000248>
- Shannon CE (1948) A Mathematical Theory of Communication. *Bell System Technical Journal* **27**(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shao Z, Janse E, Visser K and Meyer AS (2014) What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology* **5**. <https://doi.org/10.3389/fpsyg.2014.00772>
- Surrain SH and Luk G (2017) Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*. <https://doi.org/10.1017/S1366728917000682>
- Tomoschuk B, Ferreira VS and Gollan TH (2019) When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition* **22**(3), 516–536. <https://doi.org/10.1017/S1366728918000421>
- Unsworth S (2016) Early child L2 acquisition: Age or input effects? Neither, or both?. *Journal of Child Language* **43**(3), 608–634. <https://doi.org/10.1017/S030500091500080X>
- van den Noort M, Struys E, Bosch P, Jaswetz L, Perriard B, Yeo S, Barisch P, Vermeire K, Lee S-H and Lim S (2019) Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? A systematic review. *Behavioral Sciences* **9**(3), 27. <https://doi.org/10.3390/bs9030027>
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K and Yutani H (2019) Welcome to the tidyverse. *Journal of Open Source Software* **4**(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wodniecka Z, Casado A, Kalamala P, Marecka M, Timmer K and Wolna A (2020) The dynamics of language experience and how it affects language and cognition. In *Psychology of Learning and Motivation* (Vol. 72). Elsevier, pp. 235–281. <https://doi.org/10.1016/bs.plm.2020.02.005>