AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 1, 157–174
https://doi.org/10.1037/xge0001263

Bilingualism Caught in a Net: A New Approach to Understanding the Complexity of Bilingual Experience

Patrycja Kałamała[1, 2], Adam Chuderski[1], Jakub Szewczyk[3, 4], Magdalena Senderecka[1], and Zofia Wodniecka[2]

[1] Institute of Philosophy, Jagiellonian University
[2] Psychology of Language and Bilingualism Lab, Institute of Psychology, Jagiellonian University
[3] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[4] Donders Institute for Brain, Cognition and Behavior, Radboud University

The growing importance of research on bilingualism in psychology and neuroscience motivates the need for a psychometric model that can be used to understand and quantify this phenomenon. This research is the first to meet this need. We reanalyzed two data sets (N = 171 and N = 112) from relatively young adult language-unbalanced bilinguals and asked whether bilingualism is best described by the factor structure or by the network structure. The factor and network models were established on one data set and then validated on the other data set in a fully confirmatory manner. The network model provided the best fit to the data. This implies that bilingualism should be conceptualized as an emergent phenomenon arising from direct and idiosyncratic dependencies among the history of language acquisition, diverse language skills, and language-use practices. These dependencies can be reduced to neither a single universal quotient nor to some more general factors. Additional in-depth network analyses showed that the subjective perception of proficiency along with language entropy and language mixing were the most central indices of bilingualism, thus indicating that these measures can be especially sensitive to variation in the overall bilingual experience. Overall, this work highlights the great potential of psychometric network modeling to gain a more accurate description and understanding of complex (psycho)linguistic and cognitive phenomena.

Keywords: bilingualism, network modeling, latent variable, factor analysis, individual differences

In recent years, there has been an upsurge in multidisciplinary research on bilingualism. Substantial progress has been made in identifying the brain bases of bilingualism (Del Maschio et al., 2020; DeLuca et al., 2019, 2020; Hernandez et al., 2015; Pliatsikas, DeLuca, & Voits, 2020; Vaughan & Giovanello, 2010) and its relationships with various domains of human functioning, such as aging, cognitive control, and working memory (de Bruin, 2019; Leivada et al., 2020; van den Noort et al., 2019). All this evidence has translated into greater public awareness of bilingualism, in turn leading to improved language-learning programs and prevention of the marginalization of minority-language groups. However, despite the accumulating knowledge on bilingualism, there remains an open debate on the conceptualization and measurement of this phenomenon.

Bilingualism is usually defined as achieving a state of communicative knowledge of two or more languages (Grosjean & Li, 2013). Yet such a definition does not consider the richness and variety of bilingual experiences. Currently, researchers agree that bilingualism is a complex and diverse experience that should be described on several dimensions, including bilingualism onset, language proficiency, daily language use, and/or language switching (de Bruin, 2019; Leivada et al., 2020; Luk & Esposito, 2020; Marian & Hayakawa, 2021; Surrain & Luk, 2017). Each of these dimensions can be measured using different indices. For example, bilingualism onset can be quantified as the age of language acquisition or the age of active communication in a language. Daily language use can be represented as the proportion of time spent using a language or the variation in the use of different languages (so-called language entropy; Gullifer & Titone, 2019; Kałamała et al., 2020). Examining the relations between different indices of bilingualism is essential to determine the extent to which different measures reflect individual

This article was published Online First August 1, 2022.

Patrycja Kałamała https://orcid.org/0000-0003-0045-1832

Upon acceptance of the article, all data and scripts will be available at the Open Science Framework: https://osf.io/epf4y/. The data in this article were presented at the 2021 Annual Meeting of the Psychonomic Society. We have no conflicts of interest to disclose. The research was funded by a National Science Center grant (2017/27/N/HS6/01029 to Zofia Wodniecka). During work on the article, Patrycja Kałamała was supported by National Science Center grants (2017/27/N/HS6/01029; 2020/36/T/HS6/00363) and the Foundation for Polish Science (START); Magdalena Senderecka was supported by a National Science Center grant (2015/19/B/HS6/00341). The authors thank Joanna Durlik for providing the data (Data Set 2) and Michael Timberlake for proofreading.

Correspondence concerning this article should be addressed to Patrycja Kałamała, Institute of Philosophy, Jagiellonian University, Grodzka 52, 33–332 Cracow, Poland, or Zofia Wodniecka, Psychology of Language and Bilingualism Lab, Institute of Psychology, Jagiellonian University, Ingardena 6, 30–060 Cracow, Poland. Email: patrycja.kalamala@uj.edu.pl or zofia.wodniecka@uj.edu.pl

differences in bilingual experience. This would clarify how bilingualism should be conceptualized while also providing a valid approach to measuring this phenomenon. However, to the best of our knowledge, there is no psychometric study that has investigated the structure of individual differences in bilingualism. Given this substantial gap in the available literature, we aimed to establish the first psychometric model of bilingualism and test its validity in a fully confirmatory framework.

While there have been a number of studies on the interactions between the indices of bilingualism and various environmental and cognitive constructs (e.g., socioeconomic status or executive functions), surprisingly few studies have investigated the relationships between the indices of bilingualism per se (Leivada et al., 2020). Some of them used exploratory factor analysis (EFA) to test the reliability and validity of questionnaires that probe language proficiency and daily language use (Anderson et al., 2018, 2020; Li et al., 2006, 2014; Luk & Bialystok, 2013; Marian et al., 2007) or language-switching practices (Rodriguez-Fornells et al., 2011). Importantly, in some studies (Anderson et al., 2018; Luk & Bialystok, 2013), factors related to language proficiency and daily language use correlated positively, thus suggesting that these dimensions are related. Further evidence for the relationships between the indices of bilingualism comes from regression studies. In these studies, language proficiency (measured via self-assessment in Gullifer & Titone, 2019, or via self-assessment and objective tests in Gullifer et al., 2020, and Kałamała et al., 2021) was associated with the age of second-language (L2) acquisition, the percentage of daily language use, and the diversity of language use (language entropy), thus suggesting that language proficiency displays multiple associations with other indices of bilingualism.

Altogether, research has shown remarkable interindividual variability in bilingual experience. These studies, however, are fragmented in the sense that they have typically accounted for only some of the indices that are considered important in defining bilingualism (de Bruin, 2019; Leivada et al., 2020). The use of regression further adds to this fragmentation as linear modeling ignores the potential (though reasonable to expect) relationships between variables (i.e., relationships between predictors are important insofar as they explain the variance of the dependent variable). In consequence, little is known about how different indices of bilingualism co-occur and collectively translate into individuals' bilingual experiences.

Psychometric Approach: The Factor Versus the Network Model of Bilingualism

The accumulating evidence on individual differences in bilingual experience motivates the need for a psychometric approach to capture and understand the complexity of bilingual experience. In fact, the need to investigate systematic variation in bilingual experiences has been recognized in many recent articles (Backer & Bortfeld, 2021; Beatty-Martínez & Titone, 2021; Blanco-Elorrieta & Caramazza, 2021; de Bruin et al., 2021; DeLuca et al., 2019; Navarro-Torres et al., 2021; Pliatsikas, DeLuca, & Voits, 2020). In particular, research posits that systematic variation in bilingual experience may give rise to a variety of bilingual phenotypes that display different patterns of relationship between language and cognitive processes (Beatty-Martínez & Titone, 2021; Navarro-Torres et al., 2021). The psychometric model of bilingualism should capture the variation in bilingual experience and therefore clarify how bilingualism should be conceptualized and operationalized in the literature. Such evidence is particularly important in order to counteract the current fragmentation of the literature and define the boundary conditions under which bilingualism can affect neural architecture and interact with other psychological/environmental phenomena (for similar arguments, see Beatty-Martínez & Titone, 2021; Marian & Hayakawa, 2021; Navarro-Torres et al., 2021).

Individual differences in psychological traits and abilities have traditionally been examined using a latent variable framework in which the variance shared by a number of manifest variables is described as a latent variable (also called a factor). Within this framework, the latent variable is assumed to represent a directly unobservable theoretical construct that generates and explains correlations among the manifest variables (Borsboom et al., 2004; Edwards & Bagozzi, 2000; Kline, 2016). When applied to bilingualism, the factor structure would imply that there are several generalizable dimensions that define the relationships between the indices of bilingualism and thereby explain bilingual experience. These dimensions could be directly correlated with each other, or they could all be explained by the same higher-order construct of bilingualism (thus constituting a hierarchical model).

The utility of the latent variable framework to quantify bilingualism has recently been considered by Marian and Hayakawa (2021). By analogy to the well-established general intelligence factor ("*g* factor"), these researchers discussed the theoretical possibility of representing individual differences in bilingual experience in the form of a general bilingualism factor (called "*B* factor" in their article). Although the authors did not flesh out this model in detail, the very possibility of reducing the complexity of bilingualism to a single generalizable latent variable seems promising. This would offer a unified framework that enables direct comparisons between different bilingual communities while providing a single and universal measure of bilingualism ("bilingualism quotient" in their article). Thus, evidence for a *B* factor would be a great contribution to the field. However, in our opinion, the latent variable framework may not provide an adequate description of the bilingualism construct and may consequently lead to an incomplete (if not inadequate) understanding of bilingualism itself. To explain our viewpoint, we will continue the comparison of bilingualism and intelligence.

As outlined above, the latent variable framework relies on the assumption that the correlating variables are manifestations of a common underlying construct. The factor model of intelligence fits perfectly with this assumption. Since intelligence is probed by means of cognitive tests that tap various specific intellectual abilities, their correlations can be interpreted as evidence of a unitary and developmentally stable intellectual ability (Deary, 2014; McGrew, 2009). The situation is different in the case of bilingualism. Contrary to intelligence, bilingualism is a heterogeneous and dynamically changing experience that depends not only on language skills (e.g., verbal fluency, vocabulary size) but also on past and current language-use experiences, for example, daily language use or language acquisition (Beatty-Martínez & Titone, 2021; DeLuca et al., 2019; Pliatsikas, DeLuca, & Voits, 2020; Surrain & Luk, 2017). In our view, the pursuit of a common basis of all the components of bilingualism seems to be misguided. For example, the daily practice of language use such as language switching is unlikely to be a manifestation of the same underlying construct as vocabulary size. Furthermore, the latent variable framework implies that theoretical constructs (in psychology, mental abilities, styles, traits, etc.) are not directly identifiable empirically but can only be probed indirectly by measuring their observable manifestations. However, past and current language-use experiences are not exemplary manifestations of higher-order constructs. Instead, they directly point to specific life circumstances without the need to refer to latent variables. Yet even if a manifest variable directly points to its construct when such a variable is included in the factor analysis, the model attempts to frame it into a factor with other variables. Although different factor solutions are possible, the typological heterogeneity of the manifest variables could lead to a situation in which the variables that index bilingualism do not form a coherent and robust factor model.

As argued above, the relationships between the indices of bilingualism may not be easily attributed to the presence of latent variables and their structure. The experiential nature of bilingualism together with the typological heterogeneity of its measures suggest that bilingualism stems from direct and complex interactions between specific language skills and diverse language-use experiences. Framing heterogeneous variables into factors may obscure or even wash out important yet qualitatively distinct facets of the bilingual experience, which is complex and diverse at the individual level. Therefore, unlike the *g* factor, which can be considered a common basis of different specific intellectual abilities, bilingualism is likely a phenomenon that emerges from complex and direct interactions between language acquisition, language skills, and language-use habits.

The idea of direct and complex interactions between the indices of bilingualism aligns with the latest advances in psychometrics, namely the psychometric network framework (Borsboom et al., 2021). This framework is derived from the network perspective of psychology, in which observable behaviors are assumed to emerge from a network of interacting psychological, environmental, and/or biological components. Psychometric network models give insight into this potentially multidimensional interplay. In contrast to the latent variable framework, which focuses on variance that is shared across all variables, network modeling highlights variance that is unique to pairs of variables (or clusters of variables when redundant measures are used; Borsboom et al., 2021; Fried & Cramer, 2017). Within the network framework, variables correlate with each other not because they indicate the same directly unobservable (latent) construct but because of direct and reciprocal interactions. The pattern of interactions can be visualized in the form of a network in which variables are represented as nodes. The presence of an edge between any two nodes implies the existence of a unique and direct (partial) correlation that persists when possible shared dependencies on other nodes in the network are controlled for. Moreover, unlike factor analysis, which partitions the variances into separable sources, the network framework provides the opportunity to better understand the unique role of each manifest variable. Here, the number, type, and strength of the relationships between individual variables provide information about their overall importance to the network, thus allowing identification of the variables that play the most central role in shaping the targeted construct (Bringmann & Eronen, 2018; Costantini et al., 2015; Opsahl et al., 2010).

The psychometric network framework is a promising alternative to the traditional factor models of psychological constructs because in this framework, the variables do not need to refer to more general factors but form a construct through a network of direct and reciprocal connections. In consequence, researchers in the past decade have begun to successfully apply the network framework to psychological phenomena that have traditionally been viewed from a latent variable perspective, such as intelligence, personality traits, and depression (for a review, see Borsboom et al., 2021). The psychometric network framework also seems to be a promising alternative to the hypothetical factor models of bilingualism. Here, the complexity of bilingual experience would not be explained by the existence of more general and separable dimensions of bilingualism; instead, the variables themselves—through direct and reciprocal interactions—would define bilingualism.

Present Study

To establish the psychometric model of bilingualism, we reanalyzed two data sets (N = 171 and N = 112) from young adult language-unbalanced bilinguals (mostly first-language [L1]-dominant bilinguals embedded in the L1 environment). The extensive data sets made it possible to compute multiple indices of bilingualism that can be grouped into the following four categories: the onset of bilingualism (the age of L2 acquisition, the age of active L2 use), L2 proficiency (self-ratings for four basic language skills, the LexTALE test, the semantic fluency test), relative language use (the percentage of daily time spent using L2 compared to other languages; the diversity of language use, also called language entropy), and language-switching behavior (language mixing, intersentential code switching, intrasentential code switching). The measures within the L2 proficiency and onset of bilingualism categories concerned L2 experience. The other measures took into account the relative contribution of languages a person uses in daily life.

As reviewed above, the psychometric structure of theoretical constructs can be approached using two different frameworks, each of which proposes a different perspective on the relationships between the variables. The factor framework explains the correlations in terms of latent variables (common variance), whereas the network framework explains these in terms of direct mutual dependencies (partial correlations). To reflect bilingualism as accurately as possible, we compared the statistical validity of both frameworks (i.e., respective statistical models) in describing the pattern of correlations among the various indices of bilingualism, as evidenced by each model's fit to the data. The factor and network structures were first established for one data set, and then they were validated using the second independent data set in a fully confirmatory manner. The stability of the network model was additionally tested by means of bootstrap.

Since the factor and network frameworks assume contrasting data-generating mechanisms, they consequently lead to different substantive interpretations of the targeted construct. A relatively better fit of the factor model would imply that there are one or several generalizable dimensions of bilingual experience that determine the relationships between the specific indices of bilingualism and thereby explain a major part of bilingual experience (in that case, the complexity of bilingualism would only be apparent). In contrast, if the network model yielded a better fit, then the various indices of bilingualism could be considered as actively cocreating bilingual experience: Their truly direct and idiosyncratic interactions would lead to the emergence of bilingualism with no need to rely on any higher-order constructs. Consequently, this work has considerable potential to unravel the basic properties of the bilingualism phenomenon and substantially contribute to the way it should be conceptualized in the literature. On a methodological level, the research will show whether or not the current quests for a bilingualism quotient (Marian & Hayakawa, 2021; see also Backer & Bortfeld, 2021; Beatty-Martínez & Titone, 2021; Navarro-Torres et al., 2021) are likely to be successful.

The superiority of the hierarchical factor model over the network model would indicate the possibility of extracting a single quantifiable index of bilingualism. Alternatively, in-depth psychometric network analysis can show which variables are the most central in the network. The higher the centrality of a measure, the greater its contribution to network connections. Consequently, the so-called centrality analysis can indicate which variables have the greatest potential to reflect overall variability in bilingual experience (more central variables) and which indicate more unique aspects of bilingual experience (less central variables).

Method

Participants

Data Set 1 included 171 participants (*M* age 24.0 years, *SD* = 4.6; 132 women); Data Set 2 included 112 participants (*M* age 28.0 years, *SD* = 6.0; 82 women). Table 1 presents the data concerning the participants' language characteristics. In both data sets, the participants were relatively young adult language-unbalanced Polish-English bilinguals. All of them were raised in Poland and acquired Polish (L1) in early childhood (before the age of 4). Seventeen participants of Data Set 1 and 10 participants of Data Set 2 also acquired English (L2) in early childhood, while the others started learning L2 in elementary school. On average, the participants started using L2 more intensively when they attended junior high school. Participants from both data sets considered their L2 proficiency as intermediate to high and scored relatively highly in a vocabulary test for advanced learners of English (LexTALE; Lemhöfer & Broersma, 2012). On average, they used their L2 for half of the day and moderately often used more than one language in daily communication (as indicated by language entropy, language mixing, intrasentential code switching, and intersentential code switching). In addition, around 28% of participants in each data set were learning additional languages at the time of testing (predominantly German, Spanish, or French). Yet the overall proficiency of these additional languages was poor, and their daily use was marginal (see Table 1). All participants in Data Set 1 were living in Poland at the time of testing, while 71 of the 112 participants in Data Set 2 were living in the United Kingdom (average length of residence in the United Kingdom = 6.3 years, *SD* = 2.0).

Regarding the between-group differences in participants' language experience, participant samples did not differ in the self-rated L2 proficiency and the LexTALE score (each $p > .05$). At the same time, the data sets differed in terms of bilingualism onset, relative language use, and language-switching behavior (each $p < .05$). Specifically, the participants in Data Set 1 acquired L2 earlier and more frequently used multiple languages in daily communication than the participants in Data Set 2. At the same time, the participants in Data Set 2 used L2 more often during the day than the participants in Data Set 1.

Procedure and Measures

The data sets were derived from two independent studies that were conducted in our laboratory. In each study, the participants completed the following questionnaires: the language questionnaire (based on Li et al., 2014; Marian et al., 2007) the Polish translation of the Code-Switching and Interactional Contexts Questionnaire (Hartanto & Yang, 2016, Appendix E) and the Patterns of Language Use Questionnaire (Kałamała et al., 2020, Appendix C). They also performed LexTALE (i.e., deciding whether a string of letters is an existing English word; Lemhöfer & Broersma, 2012) and the semantic fluency test (i.e., producing words that belong to a given semantic category within a given time limit; Linck et al., 2009). In each study, a sociodemographic background questionnaire and several nonlinguistic tasks were also administered, but they were beyond the scope of this report. Both studies were approved by the institutional review board at Jagiellonian University. Some analyses of Data Set 1 that are unrelated to the present research goal have been published elsewhere (Kałamała et al., 2020). Table 2 presents an overview of the variables included in the analyses.

Data Analysis and Modeling

The data were analyzed in R (R Core Team, 2021) using the following packages: bootnet (Epskamp et al., 2018), EFAtools (Steiner & Grieder, 2020), lavaan (Rosseel, 2012), NetworkToolbox

Table 1
Participants' Language Characteristics

	Polish (L1)		English (L2)		Additional languages	
	Data set 1 (*N* = 171)	Data set 2 (*N* = 112)	Data set 1 (*N* = 171)	Data set 2 (*N* = 112)	Data set 1 (*N* = 47)	Data set 2 (*N* = 32)
Variable	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Age of acquisition[a]	0.08 (0.41)	0.06 (0.24)	6.72 (3.41)	8.26 (4.11)	14.07 (6.22)	13.92 (3.98)
Age of active use[a]	1.07 (1.82)	0.94 (1.49)	11.99 (5.16)	14.70 (5.61)	17.68 (6.27)	16.40 (6.25)
Overall self-rated proficiency[b]	8.98 (0.12)	8.78 (0.64)	7.84 (0.91)	7.73 (1.04)	5.38 (1.75)	3.30 (1.60)
LexTALE score	—	—	79 (9)	78 (11)	—	—
Semantic fluency score	—	—	10.62 (5.05)	19.13 (5.07)	—	—
% of daily use	56 (16)	47 (32)	41 (15)	51 (32)	10 (6)	8 (7)
Language entropy[c]	0.81 (0.28)	0.48 (0.29)				
Language mixing[d]	4.39 (2.03)	2.59 (1.56)				
Intersentential code switching[d]	4.14 (1.82)	3.00 (1.84)				
Intrasentential code switching[d]	4.83 (2.11)	3.79 (1.88)				

Note. For a detailed description of variables, see Procedure and Measures. L1 = first language; L2 = second language.
[a] Age in years. [b] The self-ratings were 1 = *no knowledge of a given language* to 9 = *native-like proficiency*. [c] Language entropy ranged from 0 to 1.59. [d] The self-ratings were 1 = *never* to 9 = *always*.

Table 2
Overview of the Variables

Instrument name (a reference to a detailed description)	Variable name	Variable operationalization
LexTALE (Lemhöfer & Broersma, 2012)	Knowledge of L2 vocabulary	The percentage of correctly identified English words
Semantic fluency[a] (for Data Set 1, see Abrahams et al., 1997, 2000; for Data Set 2, see Linck et al., 2009)	Semantic L2 fluency	The number of correctly produced words (without repetitions and proper names); for Data Set 1, the score was additionally divided by the time spent writing a single word in order to accommodate individual variation in writing speed
Language questionnaire (Li et al., 2014; Marian et al., 2007)	Age of L2 acquisition[b]	The age at which L2 was acquired
	Age of active L2 use[b]	The age at which L2 started to be used in communication
	Subjective perception of L2 skills (self-rated L2 proficiency)[b]	The average of self-ratings for listening, reading, speaking, and writing in L2
Patterns of Language Use Questionnaire (Kałamała et al., 2020)	Relative L2 use[b]	The percentage of daily time spent using L2
	Diversity of language use (language entropy)[b]	The average of language entropies for four social settings[c], weighted by time spent using languages in these settings (language entropy for each setting was computed based on the probability of the use of languages in this setting); a higher score indicates more diverse use of languages during a typical day
	Frequency of language mixing[b]	The average of self-rated language mixing for four social settings[c], weighted by time spent using languages in these settings; a higher score indicates more frequent mixing of languages within utterances during a typical day
Code-Switching and Interactional Contexts Questionnaire (Hartanto & Yang, 2016)	Frequency of intersentential code switching[b]	The average of self-rated intersentential code switching for four social settings[c], weighted by time spent using languages in these settings; a higher score indicates more frequent switching between languages between single sentences during a typical day
Code-Switching and Interactional Contexts Questionnaire (Hartanto & Yang, 2016)	Frequency of intrasentential code switching[b]	The average of self-rated intrasentential code switching for four social settings[c], weighted by time spent on using languages in these settings; a higher score indicates more frequent switching between languages within single sentences during a typical day

Note. L2 = second language.
[a] Different versions of the semantic fluency task were used in the data sets. Data Set 1 included a computerized version, i.e., words belonging to a specific semantic category written on the computer within a 2-min time limit (generation condition) and then rewritten without any time limit. Data Set 2 included a verbal version (words produced verbally; only 30-s generation condition). Both versions included the same categories (i.e., fruits and vegetables, animals, parts of the body), which were counterbalanced across participants. Since the computerized and verbal versions have been shown to strongly correlate with each other ($r = .72$; $p < .01$ for $N = 101$; derived from Rodríguez-Aranda, 2003), they were assumed to provide comparable information on semantic fluency. [b] A variable derived from a self-assessment instrument. [c] There are four types of social settings: home, work, school, and free time. For each setting, participants declare how many hours per day they typically use each of their acquired languages. If they use more than one language in a setting, they additionally assess how often they mix and switch between languages. If a responder does not spend time in a given setting, they do not complete the respective section. The division into contexts is applied to provide a more accurate and reliable representation of typical language use (for a detailed description, see Kałamała et al., 2020).

(Christensen, 2018), psych (Revelle, 2021), Psychonetrics (Epskamp, 2021), and qgraph (Epskamp et al., 2012). Prior to the psychometric analysis, the variables were centered and scaled in order to ensure a common measurement scale. There were no missing data. Upon acceptance of the article, the materials, data, and R scripts will be available at https://osf.io/epf4y.

Since reliable evidence on the psychometric structure of bilingualism is lacking, it was established in an exploratory manner based on Data Set 1. In order to verify whether the data were suitable for psychometric analysis, the Kaiser-Meyer-Olkin (KMO) test was used (Dziuban & Shirkey, 1974; Kaiser & Rice, 1974). This statistic is the ratio of the sum of the (squared) zero-order correlations to the total sum of the (squared) zero-order correlations sum and the (squared) partial correlations sum.[1] The KMO scores range between 0 (the zero-order correlations sum tends to 0) and 1 (the partial correlations sum tends to 0). A low KMO value indicates that a given variable displays idiosyncratic relationships with the other variables (indexed by substantial partial correlations) and therefore is less likely to share variance with other variable(s); a high KMO value indicates that the respective variable shares variance with other variable(s) and therefore should contribute to the factor structure. According to the guidelines (Tabachnick et al., 2019), KMO $> .50$ is required for factor analysis because a variable can only contribute to a factor structure if it shares variance with other variables (the KMO score $> .70$ is considered very good). In contrast, for the network analysis (which is based on the partial correlations), the lower the KMO score, the higher the chance that a respective variable would be included in the network model.[2]

[1] The zero-order correlation refers to the overall correlation between any two variables (and thereby suggests a possible amount of variance shared with the other variables). In turn, partial correlation reflects the correlation after extracting the part of correlation that can be attributed to any other variable.

[2] It should be noted that the KMO test does not determine which framework (factor vs. network) is more effective in describing the data because it does not assume any specific data structure and does not account for measurement error or spurious correlations. This test can only be considered as an indicator of whether a given type of analysis can be applied to the data. If a variable correlates with other variables after partial correlations are removed, it may display common variance in a factor analysis. If a variable displays partial correlations with other variables, this variable may be included in the network model. The inclusion of a variable in a model depends on the data set, the model specification procedure, and measurement error.

The exploratory factor model was established using promax rotation and maximum likelihood estimation. Only variables with KMO scores > .50 were included in the factor analysis. The number of factors was decided on the basis of the scree plot, cumulative variance explained, interpretability, and Kaiser's criterion (Floyd & Widaman, 1995). The adjacency matrix of factor loadings > .30 was referred to as "the measurement factor model." In addition to the measurement factor model, we also fit a hierarchical factor model in which a second-order factor, called "general bilingualism," was expected to explain the variance-covariance structure among the factors in the measurement factor model (for model visualizations, see Figure 1). In the factor visualization, the lines connecting latent variables (ovals) represented the correlations between the latent variables, whereas the lines from the latent variables to the other latent variables or manifest variables (squares) represented the factor loadings.

In the network analysis, we followed the guidelines by Kan and colleagues (Kan et al., 2019, 2020; see also Borsboom et al., 2021; McFarland, 2020; Schmank et al., 2019, 2021). First, the full partial correlation matrix was derived from the zero-order correlation matrix. In such a matrix, no correlation between the two variables could be attributed to any of the other variables in the network model (see also Footnote 2). In order to account for potentially spurious (false positive) correlations, the matrix was then recursively pruned using the extended Bayesian information criterion (Foygel & Drton, 2010) at $\alpha = .05$. The nodes in the resulting network represented single variables, and the edges between the nodes represented partial correlation coefficients after the pruning procedure. The adjacency matrix of the pruned correlation matrix was referred to as "the network model." In the network visualization, nodes (depicted as circles) represented variables, whereas edges (lines) represented partial correlation coefficients. The layout of the visualization was based on Fruchterman and Reingold's (1991) algorithm, which places the variables (nodes) with the largest sum of the absolute values of the partial correlations (absolute strongest connections) in the center and more strongly correlated variables closer to each other.

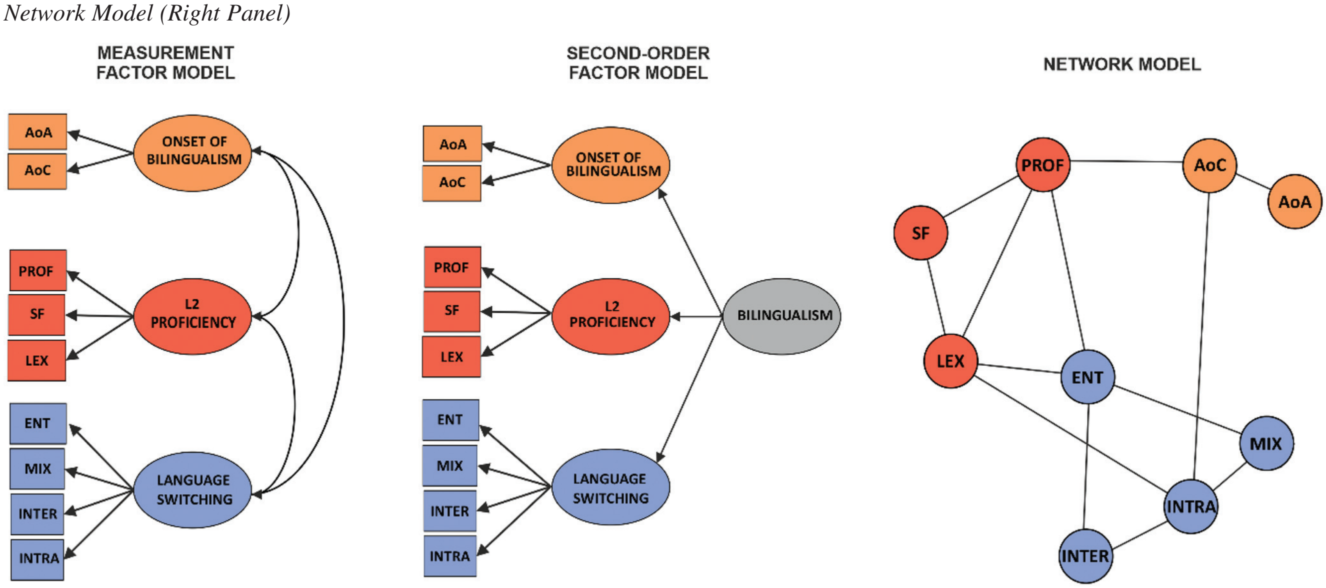
Model fitting and comparison were performed according to the procedure provided by Kan and colleagues (2019, 2020). The respective factor and network models were always fit to one and the same zero-order correlation matrix, which served as a saturated model for the model comparison. In order to verify whether the factor model and network model were correctly specified, they were first refit back to Data Set 1 (maximum likelihood estimation; factors free to correlate). Next, to verify whether the models established on Data Set 1 could be replicated in a new data set, they were fit in a fully confirmatory manner to Data Set 2.

Each resulting model was evaluated according to the standard criteria of the goodness of fit (Kline, 2016). The root mean square error of approximation (RMSEA) ≤ .05, standardized root mean square residual (SRMR) ≤ .08, and Tucker–Lewis index (TLI) and comparative fit index (CFI) > .95 were all considered to indicate a very good fit. RMSEA between .05 and .08 indicated a good fit, and RMSEA between .08 and .10 indicated a mediocre fit. TLI and CFI between .90 and .95 were considered acceptable.

In order to compare the fit across the models, we used Akaike's information criterion (AIC) and Bayesian information criterion (BIC). Both AIC and BIC adjust the model's absolute deviance

Figure 1

Graphical Representation of the Measurement Factor Model (Left Panel), the Hierarchical Factor Model (Middle Panel), and the Network Model (Right Panel)



Note. For the factor models, the lines connecting latent variables (ovals) represent the correlations between the latent variables; the lines from the latent variables to the latent and manifest variables (squares) represent the factor loadings. For the network model, the lines (edges) represent partial correlation coefficients between the manifest variables (circles). Colors of the variables correspond to the outcomes of the exploratory factor analysis. AoA = age of second language (L2) acquisition; AoC = age of active L2 use; prof = self-rated L2 proficiency; lex = knowledge of L2 vocabulary; SF = semantic L2 fluency; ent = language entropy; mix = language mixing; inter = intersentential code switching; intra = intrasentential code switching. See the online article for the color version of this figure.

from data by penalizing the number of its free parameters. This allows selection of the best-fitting but at the same time the most parsimonious model from a number of candidate models (a lower AIC/BIC value represents a better fit). BIC penalizes the free parameters more strongly than AIC does. A difference of 2 units or more in AIC/BIC indicates a significantly better fit (difference between 6 and 10 is considered strong; difference greater than 10 is considered very strong; Burnham & Anderson, 2004).

To preview the findings, the network model provided satisfactory fit to the data. Therefore, we followed with more in-depth psychometric network analysis. First, we assessed the accuracy and robustness of all edge weights (partial correlation coefficients). To this end, we used 1,000 bootstraps to estimate 95% confidence intervals (CIs) around each possible edge weight. Then, we assessed the centrality of the nodes in the network using the classic centrality indices: node strength (the sum of absolute weights of the direct connections of a node), betweenness (the probability that a node lies on the shortest path connecting any two other nodes), and closeness (the inverse of the average distance between a node and all other nodes; for mathematical definitions, see Martin & Niemeyer, 2019). The centrality indices convey how strongly a given variable is conditionally associated with other variables in the network. Node strength indicates how strongly a variable is directly associated with other variables. Betweenness reflects how often a variable acts as a bridge (mediate relationship) between any two other variables. Closeness informs how much a variable is affected by changes in any part of the network, and vice versa. Centrality estimates were z standardized ($M = 0$, $SD = 1$). The most central variables were those that consistently fell above the average (z score > 0) across the centrality indices.

The stability of the centrality estimates was evaluated using a subsetting bootstrap (i.e., by dropping participants and reestimating the network in 1,000 attempts). If the order of the nodes from a network in which some participants were dropped is highly correlated with the order of the nodes from the original network, the index can be considered stable (Epskamp et al., 2018). To quantify the stability of the centrality indices, we computed the centrality-stability (CS) coefficients. These inform about the maximum proportion of cases/participants that can be dropped while still retaining 95% probability that the correlation between the centrality based on the entire sample and that of the bootstrapped subsamples is at least .70 (representing a very large effect). The CS coefficient should be at least .25 and preferably greater than .50 for the centrality index to be stable (Epskamp et al., 2018).

Results

Descriptive Statistics and Correlations

Descriptive statistics and the zero-order correlation matrices for the data sets are shown in Table 3. The three variables with demonstrated reliability estimated satisfactory internal consistency in both data sets.[3] Most of the variables were normally distributed (i.e., their absolute values of skewness and kurtosis were lower than 2, which is considered acceptable; Kline, 2016). The variables displayed substantial variability, as indicated by the standard deviations and the minimum-maximum value ranges. The data sets were comparable in terms of the magnitude and the direction of the correlations. In both data sets, the age of L2 acquisition and the age of active L2 use showed a strong intercorrelation. The language-switching variables (language mixing, intersentential code switching, intrasentential code switching) also demonstrated strong intercorrelations. Correlations among the variables indicating L2 proficiency (self-rated L2 proficiency, knowledge of L2 vocabulary, semantic L2 fluency) were moderate in both data sets. Language entropy correlated most strongly with the variables pointing to language switching, while relative L2 use correlated most strongly with knowledge of L2 vocabulary. Overall, the results suggest that different variables of bilingualism yield moderate-to-strong intercorrelations in both data sets.

Exploratory Analysis

Table 4 shows the outcomes of the KMO test and the EFA. Overall, the KMO test suggested that Data Set 1 was moderately well suited for factor analysis. Since relative L2 use fell below the .50 threshold in the KMO test, this variable was excluded from the EFA. After its exclusion, the other variables were sufficiently intercorrelated for the EFA requirements (overall KMO score = .72; no variable fell below KMO = .50). The EFA demonstrated that the three-factor solution best described the data, $\chi^2(12) = 20.12$, $p = .065$, RMSEA [90% CI] = .06 [.00, .11] (for factor loadings, see Table 4). Correlations among the factors were weak to moderate ($r = -.37$ for Factors 1 and 2; $r = .21$ for Factors 1 and 3; $r = -.18$ for Factors 2 and 3). The model explained 56% of the total variance, but two of the three factors only accounted for a small portion of it (see Table 4). The variables with factor loadings greater than .30 constituted the measurement factor model. Language entropy, language mixing, intrasentential code switching, and intersentential code switching together reflected language-switching behavior; age of L2 acquisition and age of active L2 use jointly indicated the onset of bilingualism; and self-rated L2 proficiency along with semantic L2 fluency and knowledge of L2 vocabulary reflected L2 proficiency.[4]

[3] Reliability of the age of L2 acquisition, age of active L2 use, and semantic L2 fluency could not be assessed as these variables consisted of single measurements. Reliability of relative L2 use, language entropy, intersentential code switching, and intrasentential code switching would be meaningless as differences in self-assessment across the social settings were desirable. The adequate reliability of the semantic fluency task has been shown in previous work (Cohen & Stanczak, 2000; van den Berg et al., 2017; Woods et al., 2016). Since the other variables refer to relatively objective facts, they are largely assumed to be credible (de Bruin, 2019; Leivada et al., 2020; Marian & Hayakawa, 2021). Their credibility was also supported by an additional analysis performed on a longitudinal data set collected in our laboratory (63 Polish-English bilinguals; a 7-month period between two subsequent testing sessions). The intraclass correlation coefficients (internal consistency for a two-way fixed model) for the age of L2 acquisition and the age of active communication in L2 were 0.91 and 0.81, respectively, which indicates satisfactory reliability of the measurements (Cicchetti, 2001).

[4] To provide a full picture of the analyses, we also carried out factor analyses with relative L2 use included. The exploratory factor model including this variable poorly fit the data, $\chi^2(18) = 36.50$, $p = .006$, RMSEA [90% CI] = 0.08 [0.04, 0.11], and relative L2 use did not load any factor uniquely. Confirmatory factor models including this variable did not converge.

Table 3
Correlation Matrix and Descriptive Statistics for Data Set 1 (for Data Set 2)

Variable	AoA	AoC	Prof	Lex	SF	Use	Ent	Mix	Inter	Intra
AoA	1	**0.65 (0.66)**	**−0.26 (−0.24)**	−0.01 (**−0.20**)	−0.06 (0.00)	**0.23** (0.04)	−0.07 (−0.05)	**−0.17** (−0.03)	**−0.15** (−0.03)	**−0.22** (−0.06)
AoC		1	**−0.29 (−0.28)**	−0.01 (−0.02)	−0.07 (0.04)	**0.19** (0.13)	**−0.18** (−0.13)	**−0.32** (−0.07)	**−0.29** (−0.07)	**−0.32** (−0.14)
Prof			1	**0.33 (0.51)**	**0.35 (0.34)**	0.08 (**0.50**)	**0.17** (0.10)	**0.20** (−0.05)	**0.15** (0.07)	**0.28** (0.03)
Lex				1	**0.46 (0.21)**	**0.32 (0.39)**	−0.14 (0.08)	0.06 (0.06)	0.04 (−0.01)	**0.15** (−0.04)
SF					1	0.02 (**0.23**)	0.01 (−0.01)	0.07 (−0.07)	0.09 (−0.02)	0.11 (−0.17)
Use						1	−0.02 (−0.08)	−0.06 (−0.02)	0.09 (−0.05)	0.05 (−0.09)
Ent							1	**0.46 (0.42)**	**0.43 (0.43)**	**0.42 (0.27)**
Mix								1	**0.64 (0.53)**	**0.80 (0.58)**
Inter									1	**0.69 (0.62)**
Intra										1
M	6.72 (8.26)	11.99 (14.70)	7.84 (7.73)	79 (78)	10.62 (19.13)	41 (51)	0.81 (0.48)	4.39 (2.59)	4.14 (3.00)	4.83 (3.79)
SD	3.41 (4.11)	5.16 (5.61)	0.91 (1.04)	9 (11)	5.05 (5.07)	15 (32)	0.28 (0.29)	2.03 (1.56)	1.82 (1.84)	2.11 (1.88)
Min	0 (0)	0 (3)	3 (5)	44 (51)	2.58 (7)	11 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Max	26 (22)	35 (28)	9 (9)	98 (98)	33.68 (34)	83 (100)	1.59 (1.17)	8.62 (6.90)	9 (8.25)	9 (8)
Skewness	1.78 (1.07)	0.85 (0.23)	−1.17 (−0.47)	−0.43 (−0.13)	1.40 (0.25)	0.47 (0.11)	0.13 (0.26)	0.04 (0.82)	0.02 (0.85)	−0.17 (0.34)
Kurtosis	6.81 (1.35)	2.05 (−0.43)	3.64 (−0.73)	0.24 (−0.68)	2.99 (0.02)	−0.12 (−1.42)	0.22 (−0.55)	−0.95 (0.18)	−0.70 (0.08)	−0.94 (−0.51)
Reliability[a]	—	—	0.90 (0.91)	0.85 (0.89)	—	—	—	0.85 (0.85)	—	—

Note. AoA = age of L2 acquisition; AoC = age of active L2 use; prof = self-rated L2 proficiency; lex = knowledge of L2 vocabulary; SF = semantic L2 fluency; use = relative L2 use; ent = language entropy; mix = language mixing; inter = intersentential code switching; intra = intrasentential code switching; min = minimum; max = maximum. $p < .05$ bolded.
[a] For prof and mix, the standardized Cronbach's α was computed; for lex, the split-half correlations between odd and even trials were computed and adjusted using the Spearman-Brown prophecy formula (for details, see Kałamała et al., 2020).

To establish the network model, we followed the network analysis procedure described in Data Analysis and Modeling. As indicated above, relative L2 use was excluded from the confirmatory factor model because its KMO was too low and it hindered the calculation of the model (see also Footnote 4). Since the network analysis is based on partial correlation, this variable could, in principle, be included in the network model. However, comparison of the models' fit using AIC and BIC is only possible if the models are derived from exactly the same zero-order correlation matrix (Burnham & Anderson, 2004). Therefore, to ensure comparability, relative L2 use was also deliberately excluded from the current network analysis. Figure 1 depicts the measurement factor model, the hierarchical factor model, and the network model, all of which were exploratorily established on Data Set 1.

Confirmatory Analysis and Model Comparison

First, the measurement factor, the hierarchical factor, and the network models were fit to Data Set 1 to verify whether they were correctly specified. The measurement factor model with no constraints was inadmissible (the age of active L2 use produced a factor loading greater than 1.0 and a negative error variance). In order to adjust the model, the loadings of the age of L2 acquisition and age of active L2 use were constrained to be equal in both the measurement and the hierarchical factor models. Table 5 presents the fit statistics. The measurement factor model and the hierarchical factor model were comparable in terms of fit and both moderately fit the data. The network model fit the data very well (all the connections were identified). The comparison of AIC and BIC further indicated that the network model outperformed the measurement factor model ($\Delta_{AIC} = 44.57$, $\Delta_{BIC} = 66.56$) as well as the hierarchical factor model ($\Delta_{AIC} = 44.61$, $\Delta_{BIC} = 66.60$). The parameters of the models can be found in Tables A1–A3 of Appendix A.

The established models were subsequently fit in a truly confirmatory way to Data Set 2 (for statistics, see Table 5). As in Data Set 1, the loadings of the age of L2 acquisition and age of active L2 use were constrained to be equal in both the measurement factor model and the hierarchical factor model. In terms of the absolute fit, the factor models and the network model provided acceptable fits and thus replicated in Data Set 2. At the same time, the network model fit the data better than any of the factor models (for the measurement model, $\Delta_{AIC} = 18.96$, $\Delta_{BIC} = 37.99$). Since the network model provided the most precise description of the data in two independent data sets, it can be concluded that the network framework is better suited to studying individual differences related to bilingualism than the factor framework.

Psychometric Network Analysis on the Combined Data Set

Since relative L2 use was nonfactorable in Data Set 1 (see Exploratory Analysis), this variable did not contribute to the analyses reported above. However, the lack of an unequivocal contribution to the factor structure does not necessitate the exclusion of this variable from the network model. In contrast, a satisfactory fit of the network model to the entire variable set would serve as an important argument in favor of the network framework; it would also provide a broader and more comprehensive description of bilingual experience. Because the network model derived from Data

Table 4
Kaiser-Meyer-Olkin Test and Exploratory Factor Analysis for Data Set 1

Variable	KMO score	Factor 1 (Language-switching behavior)	Factor 2 (Onset of bilingualism)	Factor 3 (L2 proficiency)
Age of L2 acquisition	0.61	0.09	**0.83**	0.00
Age of active L2 use	0.69	−0.07	**0.79**	0.02
Self-rated L2 proficiency	0.76	0.07	−0.25	**0.43**
Knowledge of L2 vocabulary	0.55	−0.06	0.10	**0.82**
Semantic L2 fluency	0.57	−0.05	0.00	**0.60**
Relative L2 use	0.47			
Language entropy	0.83	**0.54**	0.00	−0.17
Language mixing	0.74	**0.88**	0.02	0.00
Intersentential code switching	0.82	**0.76**	0.00	−0.01
Intrasentential code switching	0.73	**0.89**	0.00	0.10
Overall KMO score	0.70			
Total variance captured (%)		32	12	13

Note. KMO = Kaiser-Meyer-Olkin Test; L2 = second language. Variables with bolded factor loadings constituted the measurement factor model.

Set 1 replicated in Data Set 2, the two data sets were combined in order to provide a more powerful test and thereby more precise estimation. The overall number of participants was N = 283, and most of the variables showed normal distribution (for criteria, see above). The network model established on the combined data set fit the data very well, $\chi^2(28) = 34.41$, $p = .19$, RMSEA [90% CI] = .028 [.00, .06], CFI = .99, TLI = .99. The left panel of Figure 2 presents the final network model.

To provide a full picture of the analyses, we also verified the possibility of fitting the factor model to the combined data set (N = 283). The KMO test indicated that the data were suited to a factor analysis (overall KMO = .72). The EFA demonstrated a similar factor structure as described in Confirmatory Analysis and Model Comparison (relative L2 use contributed to the L2 proficiency factor). The measurement factor model poorly fit the data, $\chi^2(32) = 147.38$, $p < .001$, RMSEA [90% CI] = .11 [.00, .13], CFI = .88, TLI = .79, which further corroborates the fact that the factor framework is not well suited to quantifying bilingualism. Additionally, we also checked whether the use of relative indices (instead of the absolute values of the indices) affects the statistics. To this end, the absolute values of age of L2 acquisition, age of active L2 use, and self-rated L2 proficiency were replaced with their relativized counterparts (e.g., self-rated L2 proficiency minus self-rated L1 proficiency). The models' fit and coefficients were largely the same as reported above. The outcomes of this additional analysis can be found on the Open Science Framework platform (link above).

The bootstrap analysis showed that the magnitudes of the connections (i.e., edge weights) were estimated accurately, which confirmed the credibility of the established network (for details, see Appendix B). The strongest connections (i.e., partial correlations r ranging from .40 to .60) emerged among variables related to language switching (language mixing, intrasentential code switching, and intersentential code switching) and variables that pointed to the onset of bilingualism (the age of L2 acquisition and the age of active L2 use). As for the variables related to L2 proficiency, self-rated L2 proficiency was moderately connected with both knowledge of L2 vocabulary and semantic L2 fluency (r = .32 and r = .28, respectively). The latter two variables of L2 proficiency did not show any direct connection between each other. Both language entropy and relative L2 use revealed multiple moderate connections with the other variables ($|r|$ ranging from .12 to .28). Language entropy demonstrated positive connections with language mixing, intersentential code switching, and self-rated L2 proficiency. At the same time, language entropy was negatively connected with semantic L2 fluency and relative L2 use. In addition to the connection with language entropy, relative L2 use was positively related to the age of active L2 use, self-rated L2 proficiency, and knowledge of L2 vocabulary.

Concerning the centrality of the variables, three of them were consistently above the average (z score > 0) across all the three centrality indices: self-rated L2 proficiency, language entropy, and age of active L2 use (see right panel of Figure 2). Language mixing also performed above the average in terms of direct connections (node

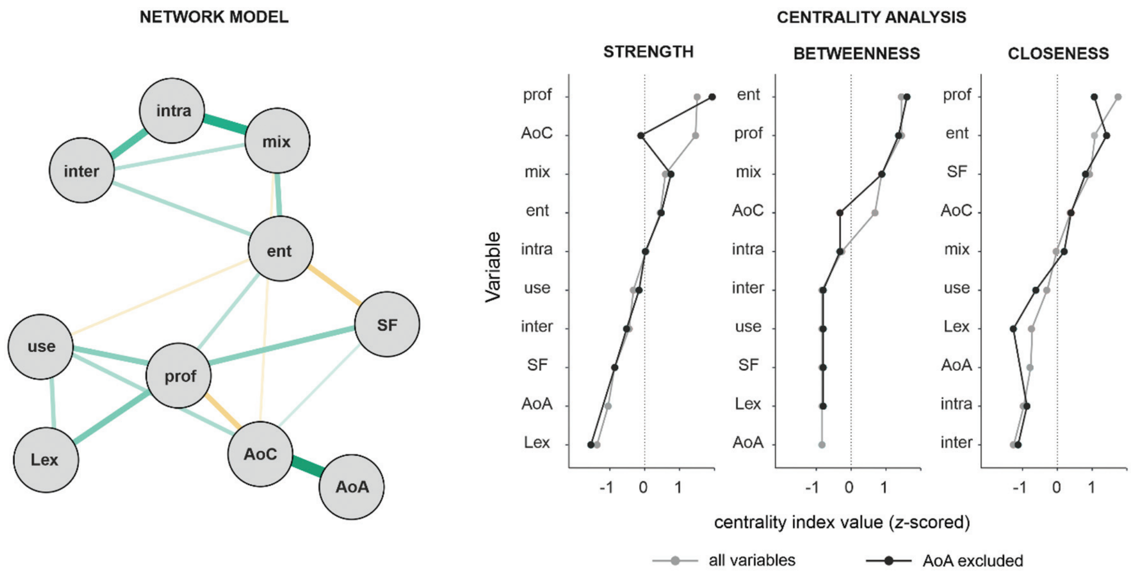
Table 5
Statistics for the Factor and Network Models for Data Set 1 and Data Set 2 (Relative L2 Use Excluded)

Data set	Model	$\chi^2(df)$	p value	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	BIC
1	Measurement	50.84 (25)	.004	0.95	0.91	0.078 [0.05, 0.11]	0.08	3,890.67	3,981.78
	Hierarchical	50.88 (25)	.002	0.95	0.91	0.078 [0.04, 0.11]	0.08	3,890.71	3,981.82
	Network	**20.27 (23)**	**.632**	**0.99**	**0.99**	**0.001 [0.00, 0.05]**	**0.03**	**3,846.10**	**3,915.22**
2	Measurement	36.00 (25)	.071	0.96	0.92	0.063 [0.00, 0.11]	0.06	2,655.21	2,734.05
	Hierarchical	36.04 (25)	.075	0.96	0.92	0.063 [0.00, 0.11]	0.06	2,655.22	2,734.06
	Network	**30.85 (23)**	**.130**	**0.97**	**0.95**	**0.055 [0.00, 0.10]**	**0.05**	**2,636.25**	**2,696.06**

Note. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root mean residual; AIC = Akaike information criterion; BIC = Bayesian information criterion. χ^2 test assumes that the factor and network models are nested with the zero-order correlation matrix (saturated model). Preferred model bolded.

Figure 2

Network Model (Left Panel) and Centrality Indices (Right Panel) for the Combined Data Set (N = 283)



Note. For the left panel, the thickness and color saturation of the edges correspond to the strength of the association between the variables (circles). Green lines represent positive partial correlation coefficients; orange lines represent negative partial correlation coefficients. For the right panel, the *y*-axis represents the variables; the *x*-axis represents the *z*-scored values of centrality indices. The gray line refers to the centrality estimates computed using all variables; the black line refers to the centrality indices computed on the variable set after exclusion of AoA. The variables are ordered by decreasing centrality for the full variable set (gray line). AoA = age of L2 acquisition; AoC = age of active L2 use; ent = language entropy; inter = intersentential code switching; intra = intrasentential code switching; lex = knowledge of L2 vocabulary; mix = language mixing; prof = self-rated L2 proficiency; SF = semantic L2 fluency; use = relative L2 use. See the online article for the color version of this figure.

strength) and the likelihood of being a bridging variable (betweenness) but not in terms of susceptibility to changes in the network (closeness; z score of $-.03$). The CS coefficient for the node strength index was .44, which indicates that the order of variables in terms of node strength was stable when varying the sample size. Conversely, the CS coefficients for betweenness and closeness were below the threshold, indicating that the exact order of the variables in terms of betweenness and closeness should be treated with some caution (for details, see Appendix B). Moreover, it should be noted that the inspection of the final network model suggested that the centrality estimates for the age of active L2 use might have been overestimated. Since the age of active L2 use mediated all dependencies of the age of L2 acquisition (see left panel of Figure 2), the strong connection between these two variables may have artificially increased the centrality indices for the age of active L2 use. Indeed, after excluding the age of L2 acquisition (a less central variable according to the centrality indices) from the network, the age of active L2 use ceased to be the central variable across all the centrality indices (z scores < 0; see right panel of Figure 2). Importantly, however, self-rated L2 proficiency, language entropy, and language mixing still remained the most central variables (z scores > 0). Collectively, the analyses indicated that self-rated L2 proficiency, language entropy, and language mixing were the most central measures of bilingual experience in this report (z scores > 0 across the centrality analyses). These variables showed the strongest direct connections with other variables (indexed by node strength); they also most often mediated the relationships between other variables (indexed by betweenness) and were most susceptible to changes in the network (indexed by closeness). In contrast, the measures related to the onset of bilingualism (especially age of L2 acquisition), inter- and intrasentential code switching, as well as the measures of linguistic knowledge (knowledge of L2 vocabulary, semantic L2 fluency) appeared to be relatively less central in the network (z scores < 0 across the centrality indices).

Discussion

This research aimed to establish the first psychometric model of bilingualism. To this end, we reanalyzed data from two independent studies in which relatively large groups of young adult language-unbalanced bilinguals completed the same set of questionnaires and tasks that probed bilingual experience. To establish a valid psychometric structure of bilingualism, we asked whether bilingualism is best described by the factor model (the number of generalizable dimensions of bilingualism that are potentially explained by a higher-order construct) or by the network model (direct and idiosyncratic pairwise dependencies between diverse language skills and language-use practices that lead to the emergence of bilingualism). Adjudicating between the factor and network models made it possible to represent bilingualism in a valid way, thus clarifying how this construct should be conceptualized and quantified in the literature.

Results Summary

The results unequivocally showed that the network model fit the two data sets better than both the measurement factor model and the hierarchical factor model. Furthermore, while not all manifest variables could be included in the factor models (relative L2 use was nonfactorable, as indicated by the KMO test), the network model fit the entire set of variables very well. The results therefore showed that the network framework was more effective than the factor framework in describing individual differences in bilingual experience.

Overall, the indices of bilingualism demonstrated moderate pairwise connections in the final network model. The centrality analysis further showed that self-rated L2 proficiency along with language entropy and language mixing were the most central variables in the network. These variables displayed the strongest direct and indirect connections with the other variables and most often mediated the relationships among other variables. The data therefore suggest that the subjective perception of L2 proficiency, diversity of language use, and language-mixing practices most prominently reflect interindividual variability and thus together may serve as the most representative indices of variation in bilingual experience. In contrast, the measures of bilingualism onset (the age of L2 acquisition, the age of active L2 use) and language skills (knowledge of L2 vocabulary, semantic L2 fluency) were found to be less central in the estimated network. This suggests that each of these measures points to a very specific aspect of bilingual experience, and therefore none of them should serve as a universal index of bilingualism. Yet while the measures of bilingualism onset and language skills do not seem to inform about the general variability in bilingual experience, their importance in research should not be disregarded. A lot of neuroscientific research has reported the effects of language acquisition or semantic fluency on language processing, cognitive functions, and brain architecture (e.g., Cargnelutti et al., 2019; Klein et al., 2014; Pliatsikas, 2019; Pliatsikas, Meteyard, et al., 2020; Wei et al., 2015). The lower centrality of these measures in the present network suggests that the effects similar to those reported in these studies could not be easily attributed to the variability in other indices of bilingualism. The centrality analysis also revealed other important properties of bilingualism and its measures. As can be seen in Figure 2 (left panel), the age of active L2 use mediated all dependencies of the age of L2 acquisition in the network, which made the age of active L2 use a more central variable. Therefore, the analyses suggest that the age of active L2 use displays more relationships with the other indices of bilingualism and thus may be a more representative index of bilingualism than the age of L2 acquisition. This finding is interesting from a methodological viewpoint because researchers typically use the latter measure to document bilingualism (de Bruin, 2019; Leivada et al., 2020; Surrain & Luk, 2017). Similarly, the frequency of language mixing derived from the Patterns of Language Use Questionnaire (Kałamała et al., 2020), which is one of the central measures in the estimated network, appears to be a more representative predictor of bilingualism than the indices of inter- and intrasentential code switching provided by Hartanto and Yang (2016). The relatively lower centrality of the measures of linguistic knowledge (i.e., knowledge of L2 vocabulary and semantic L2 fluency) might, in turn, be a consequence of the fact that these variables were the only two objective measures in the set. We will come back to this issue in Limitations and Future Directions. Moreover, it should be noted that some connections in the network may be considered counterintuitive. Although earlier age of active L2 use and greater diversity of language use were associated with better self-perception of L2 abilities, both were related to poorer semantic fluency. Therefore, the network model of bilingualism does not demonstrate a positive manifold in which all measures are positively intercorrelated (a well-established phenomenon in the psychological study of intelligence; Spearman, 1904). When combined, different indices of bilingualism form a system of multiple idiosyncratic dependencies that together explain variability in individual bilingual experiences.

Methodological and Theoretical Implications

While recent calls to account for individual variability in the bilingual experience have already begun to propel the field forward, a lingering issue is how bilingualism should be conceptualized and quantified. The present study is the first to provide a psychometric model of bilingualism. In consequence, the findings have important methodological and theoretical implications.

On the methodological level, the report provides rather strong arguments against applying the factor framework to study the psychometric structure of bilingualism. Our analyses show that the indices of bilingualism formed a complex pattern of relationships that could not be effectively reduced to a single quantitative variable. Therefore, the report suggests that the recent quests for a general bilingualism factor (Marian & Hayakawa, 2021; see also Backer & Bortfeld, 2021; Beatty-Martínez & Titone, 2021; Navarro-Torres et al., 2021) are likely a search for the holy grail. Yet the lack of convincing arguments in favor of a robust factor model does not imply that bilingualism is nonquantifiable or that researchers must apply all possible measures simultaneously. The centrality indices consistently showed that self-rated L2 proficiency along with language entropy and language mixing were the most central variables in the network, thus indicating that these measures are especially sensitive to variation in the overall bilingual experience. Importantly, however, the exact set of the most central measures may differ among bilingual communities. The proposed set was established for relatively young adult language-unbalanced bilinguals who were mostly embedded in the L1 environment. We will discuss this issue in Limitations and Future Directions.

While we provide arguments against the factor model of bilingualism, it should be noted that we do not argue against the overall usefulness of factor analysis in research on bilingualism. As Marian and Hayakawa (2021) pointed out, the multiplicity of measures is a real methodological problem in the field because differences in measurements complicate direct comparisons across different bilingual communities and laboratories. By isolating common variance, the factor framework can effectively show to what extent different measures that are assumed to reflect the same specific construct (e.g., language proficiency) actually produce consistent results. From this perspective, factor analysis is highly desirable in the field because it can help to eliminate redundant measurement tools. However, what we argue here is that the factor framework is unsatisfactory when one aims to represent the complexity of bilingualism.

On a theoretical level, the lack of compelling evidence for the factor structure of bilingualism suggests that bilingualism cannot be identified with some more general psychological constructs that would underlie individual differences in bilingual experience. Rather than being a driving force, bilingualism should be understood as an emergent phenomenon arising from idiosyncratic dependencies between language skills, language acquisition history, and language-use habits. In particular, the centrality analysis suggests that bilingualism does not merely refer to knowledge of languages but constitutes a rich experience that depends largely on current language-use practices and subjective perception of language skills rather than on acquired linguistic knowledge and actual language-learning history.

Critically, the importance of this work is not limited to the domain of bilingualism; it also has important implications for psychology in general. Nowadays, psychology as a field increasingly recognizes the fact that bilingualism is ubiquitous in human experience, that is, that there are more bilinguals in the world than monolinguals (Craik & Bialystok, 2006; U.S. Census Bureau, 1999). At the same time, research on bilingualism shows that any two individuals differ substantially in their bilingual experiences (DeLuca et al., 2019; Gullifer et al., 2020; Kałamała et al., 2021). Therefore, it becomes crucial to establish an appropriate methodological framework that would adequately capture this complex experience while allowing for its effective quantification. As shown in this study, psychometric network modeling provides a solid framework for understanding the variability of bilingual experience. In addition to this, the study also has important implications for modeling other complex (psycho)linguistic and cognitive phenomena. While many theories in (psycho)linguistics and cognitive science draw from a network perspective, the application of network science to the quantitative study of cognition has so far still been limited in scope (but see, e.g., Kan et al., 2019; Schmank et al., 2019). Moreover, much of our understanding of language and cognition comes from latent variable research. However, some factor models of cognitive constructs are criticized for their poor performance and low replicability (e.g., attentional control; Karr et al., 2018; von Bastian et al., 2020). Taking bilingualism as a prime example, we presented evidence that when studying individual differences in certain cognitive and (psycho)linguistic phenomena, it is necessary to shift the focus from investigating common variance to studying direct mutual interactions between variables. Therefore, we hope that the evidence in favor of network modeling (and against the factor framework) will draw the attention of researchers in the fields of language and cognition to the possible pitfalls of studying individual differences and serve as further evidence of the enormous potential of network modeling in psychological research.

Limitations and Future Directions

The study sheds light on how to describe and conceptualize bilingualism. However, given the pioneering nature of this work, some findings are limited and require further investigation. Although the models established on a sample of Polish-English bilinguals living in the L1 environment have been replicated on Polish-English bilinguals living in a more linguistically rich environment, some caution needs to be exerted in extrapolating the results to other bilingual communities. We believe that the superiority of the network model over the factor models is due to the complexity of bilingualism and the typological heterogeneity of its measures (for a detailed argument, see the introduction). Therefore, we predict that the very good performance of the network model (and the worse performance of the factor models) will replicate across different bilingual communities. At the same time, however, we anticipate that the specific relationships between the variables observed in the network model may differ between bilingual communities embedded in different language contexts (for arguments, see, e.g., Beatty-Martínez et al., 2020; Xie & Antolovic, 2021; Zhang et al., 2021). Here, we present the network model for relatively young adult language-unbalanced bilinguals who are mostly embedded in the L1 environment, all of whom reported regular contact with L1. Another network model may be demonstrated for bilinguals who live in environments that support the use of multiple languages or environments that significantly restrict the use of L1. Therefore, the current network model should not be considered universal. Instead, it can serve as a prototype against which the network models derived from other bilingual communities can be tested. An additional advantage is that network modeling provides statistical techniques for between-group network comparisons (Borsboom et al., 2021; Forbes et al., 2017; van Borkulo et al., 2022). Future research should thus try to determine which of the variables and connections have strong predictive validity across different bilingual communities and which of them are unique for certain language contexts. Systematic psychometric research across different bilingual communities should ultimately inform about the role of language environment/context in shaping bilingual experiences.

The indices of bilingualism selected for the current study also deserve a comment. In this study, we have focused on the dimensions of bilingualism that have received the greatest attention in the literature, namely the onset of bilingualism, language proficiency, daily language use, and language-switching behavior (Luk & Esposito, 2020; Marian & Hayakawa, 2021). Each of these dimensions was probed using multiple and well-established measures (Surrain & Luk, 2017). Therefore, we believe that bilingualism was adequately represented in the study and thereby provided a solid comparison of the factor and network framework. It should be noted, however, that out of the 10 variables, only knowledge of L2 vocabulary and semantic L2 fluency were derived from objective measurements (i.e., LexTALE and semantic fluency test, respectively). Underrepresentation of the objective measures could translate into their lower centrality in the overall network. Moreover, some researchers have also pointed to the existence of other dimensions of bilingualism that we have not accounted for in this study (e.g., the manner of L2 acquisition, identification with the culture of a foreign language, sociodemographic status; Marian & Hayakawa, 2021). Therefore, future research should consider incorporating a larger set of measures and—in an ideal scenario—balancing the number of objective and subjective variables to further inform the network model of bilingualism.

In this study, we directly focused on the complexity of bilingual experience. However, bilingualism is considered not only a complex experience but also a dynamic one that changes over time (Luk & Esposito, 2020; Wodniecka et al., 2020). Since the network framework provides statistical techniques for studying the dynamics of networks within individuals (Borsboom et al., 2021; Epskamp, 2020), another interesting avenue for future research is to adopt a longitudinal approach and examine how bilingualism evolves. If the quantity and/or quality of connections within the network systematically changes, then critical points in the development of bilingualism could be identified. Importantly, firmly establishing a developmental cascade of bilingual experience could make it possible to determine which variables and connections are crucial in the development of this phenomenon and—in a broader perspective—could provide an opportunity for a viable classification of bilingual individuals (or identification of bilingual phenotypes; Beatty-Martínez & Titone, 2021), which, although very desirable, is currently missing (Backer & Bortfeld, 2021; Blanco-Elorrieta & Caramazza, 2021; de Bruin et al., 2021; Navarro-Torres et al., 2021).

Conclusions

The complexity of bilingual experience makes the study of bilingualism fascinating but also challenging. In this study, we demonstrated how to overcome this challenge. The network model was more accurate in reflecting individual differences in bilingual experience and skills than any of the factor models. Therefore, the study provides the first empirical evidence for the idea that bilingualism is an emergent phenomenon arising from complex and unique interactions among language skills, habits, and history of language acquisition. Importantly, further analyses showed that bilingualism does not merely refer to knowledge of languages but constitutes a rich experience that depends largely on current language-use practices and subjective perception of language skills rather than on acquired linguistic knowledge and actual language-learning history. Hence, self-rated L2 proficiency, along with language entropy and language mixing, seem to have the greatest potential to reflect overall variability in the bilingual experience (but this effect requires further investigation). In a broader context, the evidence in favor of network modeling (and against the factor framework) opens the field of bilingualism up to new explorations and considerations. It would be especially appealing to trace the dynamics of bilingualism within a network in order to learn how bilingualism evolves. Overall, this work presents the enormous potential of network modeling to gain a unified description and more comprehensive understanding of bilingualism and other complex cognitive and (psycho)linguistic phenomena.

Context

We represent different research domains of psychology. Some of us focus on individual differences in complex cognitive functions; others focus on bilingualism from the perspectives of cognitive psychology and neuroscience. In this work, we joined forces to learn how to quantify and conceptualize bilingualism. The evidence in favor of the network model of bilingualism (compared to the factor model) has important implications for research on bilingualism and all other areas of psychology that focus on studying complex phenomena. The outcomes form the basis of our new research program, in which we aim to test individual differences in bilingual experience at a more detailed level (between-population and longitudinal comparisons).

References

Abrahams, S., Goldstein, L. H., Al-Chalabi, A., Pickering, A., Morris, R. G., Passingham, R. E., Brooks, D. J., & Leigh, P. N. (1997). Relation between cognitive dysfunction and pseudobulbar palsy in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *62*(5), 464–472. https://doi.org/10.1136/jnnp.62.5.464

Abrahams, S., Leigh, P. N., Harvey, A., Vythelingum, G. N., Grisé, D., & Goldstein, L. H. (2000). Verbal fluency and executive dysfunction in amyotrophic lateral sclerosis (ALS). *Neuropsychologia*, *38*(6), 734–747. https://doi.org/10.1016/S0028-3932(99)00146-3

Anderson, J. A. E., Hawrylewicz, K., & Bialystok, E. (2020). Who is bilingual? Snapshots across the lifespan. *Bilingualism: Language and Cognition*, *23*(5), 929–937. https://doi.org/10.1017/S1366728918000950

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

Backer, K. C., & Bortfeld, H. (2021). Characterizing bilingual effects on cognition: The search for meaningful individual differences. *Brain Sciences*, *11*(1), Article 81. https://doi.org/10.3390/brainsci11010081

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(6), 1022–1047. https://doi.org/10.1037/xlm0000770

Beatty-Martínez, A. L., & Titone, D. A. (2021). The quest for signals in noise: Leveraging experiential variation to identify bilingual phenotypes. *Languages*, *6*(4), Article 168. https://doi.org/10.3390/languages6040168

Blanco-Elorrieta, E., & Caramazza, A. (2021). On the need for theoretically guided approaches to possible bilingual advantages: An evaluation of the potential loci in the language and executive control systems. *Neurobiology of Language*, *2*(4), 452–463. https://doi.org/10.1162/nol_a_00041

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), Article 58. https://doi.org/10.1038/s43586-021-00055-w

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J. T. W., & Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, *128*(8), 892–903. https://doi.org/10.1037/abn0000446

Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. *Psychological Review*, *125*(4), 606–615. https://doi.org/10.1037/rev0000108

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Cargnelutti, E., Tomasino, B., & Fabbro, F. (2019). Language brain representation in bilinguals with different age of appropriation and proficiency of the second language: A meta-analysis of functional imaging studies. *Frontiers in Human Neuroscience*, *13*, Article 154. https://doi.org/10.3389/fnhum.2019.00154

Christensen, A. P. (2018). NetworkToolbox: Methods and measures for brain, cognitive, and psychometric network analysis in R. *The R Journal*, *10*(2), 422–439. https://doi.org/10.32614/RJ-2018-065

Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology*, *23*(5), 695–700. https://doi.org/10.1076/jcen.23.5.695.1249

Cohen, M. J., & Stanczak, D. E. (2000). On the reliability, validity, and cognitive structure of the Thurstone Word Fluency Test. *Archives of Clinical Neuropsychology*, *15*(3), 267–279. https://doi.org/10.1093/arclin/15.3.267

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal* *of Research in Personality*, *54*, 13–29. https://doi.org/10.1016/j.jrp.2014.07.003

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*(3), 131–138. https://doi.org/10.1016/j.tics.2006.01.007

Deary, I. J. (2014). The stability of intelligence from childhood to old age. *Current Directions in Psychological Science*, *23*(4), 239–245. https://doi.org/10.1177/0963721414536905

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), Article 33. https://doi.org/10.3390/bs9030033

de Bruin, A., Dick, A. S., & Carreiras, M. (2021). Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiology of Language*, *2*(4), 433–451. https://doi.org/10.1162/nol_a_00038

Del Maschio, N., Sulpizio, S., & Abutalebi, J. (2020). Thinking outside the box: The brain-bilingualism relationship in the light of early neurobiological variability. *Brain and Language*, *211*, Article 104879. https://doi.org/10.1016/j.bandl.2020.104879

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7565–7574. https://doi.org/10.1073/pnas.1811513116

DeLuca, V., Segaert, K., Mazaheri, A., & Krott, A. (2020). Understanding bilingual brain function and structure changes? U bet! A unified bilingual experience trajectory model. *Journal of Neurolinguistics*, *56*, Article 100930. https://doi.org/10.1016/j.jneuroling.2020.100930

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361. https://doi.org/10.1037/h0036316

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. https://doi.org/10.1037/1082-989X.5.2.155

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, *85*(1), 206–231. https://doi.org/10.1007/s11336-020-09697-3

Epskamp, S. (2021). Psychonetrics: Structural equation modeling and confirmatory network analysis (R package version 0.8.1) [Computer software]. https://CRAN.R-project.org/package=psychonetrics

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Cramer, A., Waldorp, L., Schmittmann, V., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18. https://doi.org/10.18637/jss.v048.i04

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286–299. https://doi.org/10.1037/1040-3590.7.3.286

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, *126*(7), 969–988. https://doi.org/10.1037/abn0000276

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems 23* (pp. 604–612). Neural Information Processing Systems.

Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*(6), 999–1020. https://doi.org/10.1177/1745691617705892

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software, Practice & Experience*, *21*(11), 1129–1164. https://doi.org/10.1002/spe.4380211102

Grosjean, F., & Li, P. (Eds.). (2013). Bilingualism: A short introduction. *The psycholinguistics of bilingualism* (pp. 5–25). Wiley-Blackwell.

Gullifer, J. W., Kousaie, S., Gilbert, A. C., Grant, A. M., Giroud, N., Coulter, K., Klein, D., Baum, S., Phillips, N., & Titone, D. (2020). *Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency*. PsyArXiv. https://doi.org/10.31234/osf.io/gb9nd

Gullifer, J. W., & Titone, D. (2019). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*(2), 283–294. https://doi.org/10.1017/S1366728919000026

Hartanto, A., & 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

Hernandez, A. E., Greene, M. R., Vaughn, K. A., Francis, D. J., & Grigorenko, E. L. (2015). Beyond the bilingual advantage: The potential role of genes and environment on the development of cognitive control. *Journal of Neurolinguistics*, *35*, 109–119. https://doi.org/10.1016/j.jneuroling.2015.04.002

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kałamała, P., Senderecka, M., & Wodniecka, Z. (2021). On the multidimensionality of bilingualism and the unique role of language use. *Bilingualism: Language and Cognition*, *25*(3), 471–483. https://doi.org/10.1017/S1366728921001073

Kałamała, P., Szewczyk, J., Chuderski, A., Senderecka, M., & Wodniecka, Z. (2020). Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition*, *204*, Article 104373. https://doi.org/10.1016/j.cognition.2020.104373

Kan, K.-J., de Jonge, H., van der Maas, H. L. J., Levine, S. Z., & Epskamp, S. (2020). How to compare psychometric factor and network models. *Journal of Intelligence*, *8*(4), Article 35. https://doi.org/10.3390/jintelligence8040035

Kan, K.-J., van der Maas, H. L. J., & Levine, S. Z. (2019). Extending psychometric network analysis: Empirical evidence against g in favor of mutualism? *Intelligence*, *73*, 52–62. https://doi.org/10.1016/j.intell.2018.12.004

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Klein, D., Mok, K., Chen, J.-K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, *131*, 20–24. https://doi.org/10.1016/j.bandl.2013.05.014

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Leivada, E., Westergaard, M., Duñabeitia, J. A., & Rothman, J. (2020). On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed? *Bilingualism: Language and Cognition*, *24*(1), 197–210. https://doi.org/10.1017/S1366728920000358

Lemhöfer, K., & 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., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–210. https://doi.org/10.3758/BF03192770

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

Language and Cognition, *17*(3), 673–680. https://doi.org/10.1017/S1366728913000606

Linck, J. A., Kroll, J. F., & 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., & 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*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marian, V., & Hayakawa, S. (2021). Measuring bilingualism: The quest for a "bilingualism quotient". *Applied Psycholinguistics*, *42*(2, Suppl. 2), 527–548. https://doi.org/10.1017/S0142716420000533

Martin, C., & Niemeyer, P. (2019). Influence of measurement errors on networks: Estimating the robustness of centrality measures. *Network Science*, *7*(2), 180–195. https://doi.org/10.1017/nws.2019.12

McFarland, D. (2020). The effects of using partial or uncorrected correlation matrices when comparing network and latent variable models. *Journal of Intelligence*, *8*(1), Article 7. https://doi.org/10.3390/jintelligence8010007

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Navarro-Torres, C. A., Beatty-Martínez, A. L., Kroll, J. F., & Green, D. W. (2021). Research on bilingualism as discovery science. *Brain and Language*, *222*, Article 105014. https://doi.org/10.1016/j.bandl.2021.105014

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. https://doi.org/10.1016/j.socnet.2010.03.006

Pliatsikas, C. (2019). Multilingualism and brain plasticity. In J. W. Schwieter & M. Paradis (Eds.), *The handbook of the neuroscience of multilingualism* (pp. 230–251). Wiley, Ltd. https://doi.org/10.1002/9781119387725.ch11

Pliatsikas, C., DeLuca, V., & 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

Pliatsikas, C., Meteyard, L., Veríssimo, J., DeLuca, V., Shattuck, K., & Ullman, M. T. (2020). The effect of bilingualism on brain development from early childhood to young adulthood. *Brain Structure & Function*, *225*(7), 2131–2152. https://doi.org/10.1007/s00429-020-02115-5

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2021). psych: Procedures for psychological, psychometric, and personality research (R package version 2.1.3) [Computer software]. Northwestern University. https://CRAN.R-project.org/package=psych

Rodríguez-Aranda, C. (2003). Reduced writing and reading speed and age-related changes in verbal fluency tasks. *The Clinical Neuropsychologist*, *17*(2), 203–215. https://doi.org/10.1076/clin.17.2.203.16508

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*, Article 388. https://doi.org/10.3389/fpsyg.2011.00388

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schmank, C. J., Goring, S. A., Kovacs, K., & Conway, A. R. A. (2019). Psychometric network analysis of the Hungarian WAIS. *Journal of Intelligence*, *7*(3), Article 21. https://doi.org/10.3390/jintelligence7030021

Schmank, C. J., Goring, S. A., Kovacs, K., & Conway, A. R. A. (2021). Investigating the structure of intelligence using latent variable and psychometric network modeling: A commentary and reanalysis. *Journal of Intelligence*, *9*(1), Article 8. https://doi.org/10.3390/jintelligence9010008

Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Steiner, M., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, *5*(53), Article 2521. https://doi.org/10.21105/joss.02521

Surrain, S. H., & 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. https://doi.org/10.1017/S1366728917000682

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (7th ed.). Pearson.

U.S. Census Bureau. (1999). *Statistical abstract of the United States 1999*. https://www.census.gov/

van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2022). Comparing network structures on three aspects: A permutation test. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000476

van den Berg, E., Jiskoot, L. C., Grosveld, M. J. H., van Swieten, J. C., & Papma, J. M. (2017). Qualitative assessment of verbal fluency performance in frontotemporal dementia. *Dementia and Geriatric Cognitive Disorders*, *44*(1-2), 35–44. https://doi.org/10.1159/000477538

van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S.-H., & 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), Article 27. https://doi.org/10.3390/bs9030027

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging*, *25*(2), 343–355. https://doi.org/10.1037/a0017729

von Bastian, C. C., Blais, C., Brewer, G. A., Gyurkovics, M., Hedge, C., Kałamała, P., Meier, M. E., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, L. M., Conway, A. R. A., Draheim, C., Engle, R. W., Friedman, N. P., Frischkorn, G. T., Gustavson, D. E., Koch, I., . . . Wiemers, E. A. (2020). *Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations*. PsyArXiv. https://doi.org/10.31234/osf.io/x3b9k

Wei, M., Joshi, A. A., Zhang, M., Mei, L., Manis, F. R., He, Q., Beattie, R. L., Xue, G., Shattuck, D. W., Leahy, R. M., Xue, F., Houston, S. M., Chen, C., Dong, Q., & Lu, Z.-L. (2015). How age of acquisition influences brain architecture in bilinguals. *Journal of Neurolinguistics*, *36*, 35–55. https://doi.org/10.1016/j.jneuroling.2015.05.001

Wodniecka, Z., Casado, A., Kałamała, P., Marecka, M., Timmer, K., & Wolna, A. (2020). The dynamics of language experience and how it affects language and cognition. In K. D. Federmeier & H.-W. Huang (Eds.), *Psychology of learning and motivation* (Vol. 72, pp. 235–281). Elsevier. https://doi.org/10.1016/bs.plm.2020.02.005

Woods, D. L., Wyma, J. M., Herron, T. J., & Yund, E. W. (2016). Computerized analysis of verbal fluency: Normative data and the effects of repeated testing, simulated malingering, and traumatic brain injury. *PLoS ONE*, *11*(12), Article e0166439. https://doi.org/10.1371/journal.pone.0166439

Xie, Z., & Antolovic, K. (2021). Differential impacts of natural L2 immersion and intensive classroom L2 training on cognitive control. *The Quarterly Journal of Experimental Psychology*, *75*(3), 550–562. https://doi.org/10.1177/17470218211040813

Zhang, H., Diaz, M. T., Guo, T., & Kroll, J. F. (2021). Language immersion and language training: Two paths to enhanced language regulation and cognitive control. *Brain and Language*, *223*, Article 105043. https://doi.org/10.1016/j.bandl.2021.105043

Appendix A

The Outcomes of the Confirmatory Psychometric Analysis

Table A1

Parameters of the Measurement Factor Model Fit to Data Set 1

Latent variable	Manifest variable	Factor loading	Sig	Error variance (manifest variable)	Sig
		Factor loadings			
L2 proficiency	Prof	0.55	0.001	0.69	0.001
	Lex	0.63	0.001	0.59	0.001
	SF	0.67	0.001	0.55	0.001
Onset of bilingualism	AoA	0.81	0.001	0.39	0.001
	AoC	0.81	0.001	0.31	0.001
Language switching	Ent	0.49	0.001	0.75	0.001
	Mix	0.87	0.001	0.24	0.001
	Inter	0.75	0.001	0.43	0.001
	Intra	0.92	0.001	0.15	0.001

Latent variable 1	Latent variable 2	Correlation	Sig
		Factor correlations	
L2 proficiency	Onset of bilingualism	−0.20	0.054
	Language switching	0.23	0.014
Onset of bilingualism	Language switching	−0.37	0.001

Note. L2 = second language; sig = significance; AoA = age of L2 acquisition; AoC = age of active L2 use; prof = self-rated L2 proficiency; lex = knowledge of L2 vocabulary; SF = semantic L2 fluency; use = relative L2 use; ent = language entropy; mix = language mixing; inter = intersentential code switching; intra = intrasentential code switching.

Table A2

Parameters of the Hierarchical Factor Model Fit to Data Set 1

Latent variable	Manifested variable	Factor loading	Sig	Error variance (manifest variable)	Sig
		Factor loadings			
L2 proficiency	Prof	0.52	0.001	0.69	0.001
	Lex	0.59	0.001	0.59	0.001
	SF	0.63	0.001	0.55	0.001
Onset of bilingualism	AoA	0.67	0.001	0.39	0.001
	AoC	0.67	0.001	0.31	0.001
Language switching	Ent	0.37	0.001	0.75	0.001
	Mix	0.65	0.001	0.24	0.001
	Inter	0.56	0.001	0.43	0.001
	Intra	0.69	0.001	0.15	0.001
Bilingualism (hierarchical factor)	L2 proficiency	0.38	0.019		
	Onset of bilingualism	−0.67	0.024		
	Language switching	0.87	0.052		

Note. L2 = second language; sig = significance; AoA = age of L2 acquisition; AoC = age of active L2 use; prof = self-rated L2 proficiency; lex = knowledge of L2 vocabulary; SF = semantic L2 fluency; use = relative L2 use; ent = language entropy; mix = language mixing; inter = intersentential code switching; intra = intrasentential code switching.

(*Appendices continue*)

Table A3
The Adjacency Matrix of the Network Model Fit to Data Set 1

Variable	AoA	AoC	Prof	Lex	SF	Ent	Mix	Inter	Intra
AoA	0.00	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AoC	0.61	0.00	−0.18	0.00	0.00	0.00	0.00	0.00	−0.12
prof	0.00	−0.18	0.00	0.22	0.22	0.16	0.00	0.00	0.00
Lex	0.00	0.00	0.22	0.00	0.38	−0.24	0.00	0.00	0.10
SF	0.00	0.00	0.22	0.38	0.00	0.00	0.00	0.00	0.00
Ent	0.00	0.00	0.16	−0.24	0.00	0.00	0.22	0.20	0.00
Mix	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.66
Inter	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.46
Intra	0.00	−0.12	0.00	0.10	0.00	0.00	0.66	0.46	0.00

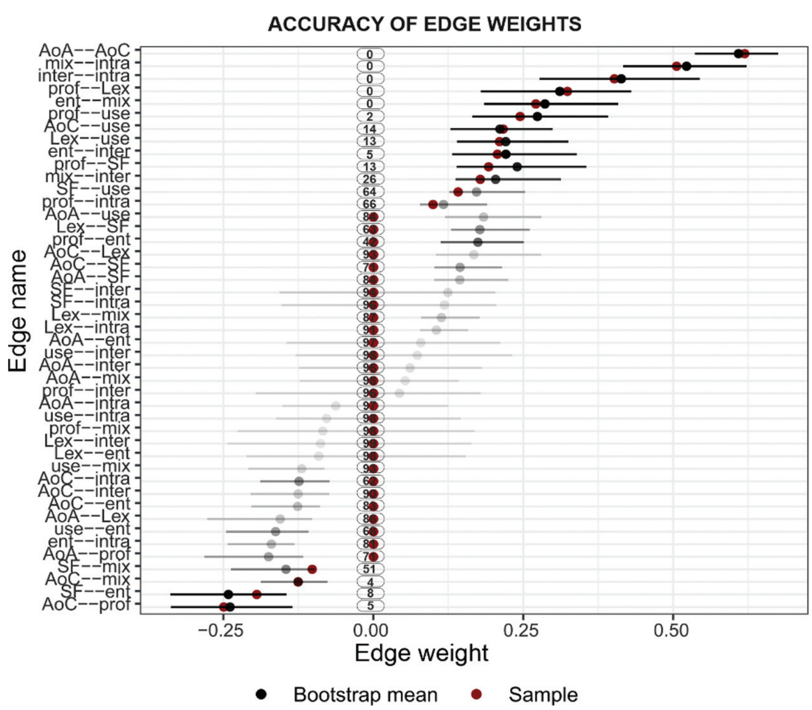
Note. AoA = age of L2 acquisition; AoC = age of active L2 use; prof = self-rated L2 proficiency; lex = knowledge of L2 vocabulary; SF = semantic L2 fluency; use = relative L2 use; ent = language entropy; mix = language mixing; inter = intersentential code switching; intra = intrasentential code switching.

Appendix B

Stability and Accuracy of the Network Model

Figure B1 presents the outcomes of the bootstrap analysis to assess the accuracy and robustness of edge weights (partial correlation coefficients). It depicts the quantile intervals only for when the parameter was not set to zero. The number in the plot indicates the percentage of bootstraps when the edge weight was estimated to be exactly zero. If the strength of the edge weight derived from the sample (i.e., red dot in the plot) is zero, then the probability of setting a parameter to zero should

Figure B1
Quantile Intervals Only for the Times the Edge Weight Was Not Set to Zero



Note. The black dots represent bootstrapped edge weights; the dark red dots represent edge weights derived from the data set. The *x*-axis represents the strength of the edge weight; the *y*-axis represents all possible edge weights. Values in the ovals indicate the percentage of 2,500 bootstraps when the edge weight was estimated to be exactly zero. AoA = age of L2 acquisition; AoC = age of active L2 use; ent = language entropy; inter = intersentential code switching; intra = intrasentential code switching; lex = knowledge of L2 vocabulary; mix = language mixing; prof = self-rated L2 proficiency; SF = semantic L2 fluency; use = relative L2 use. See the online article for the color version of this figure.

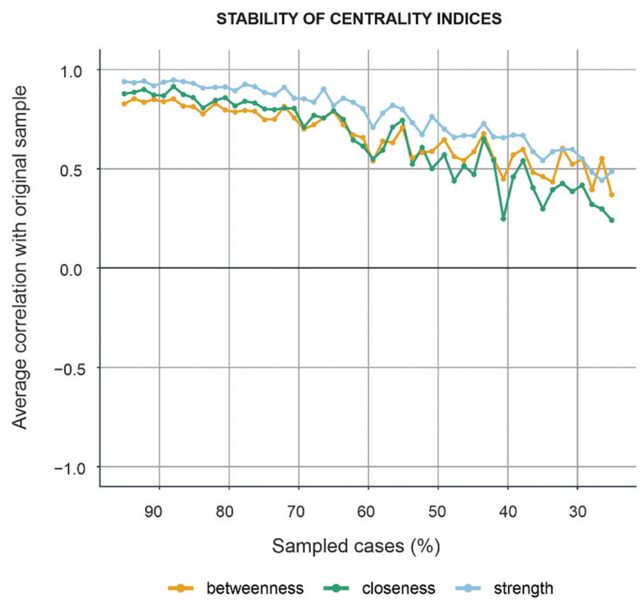
(*Appendices continue*)

be high. Otherwise, the probability of setting a parameter to zero should be low. As can be seen, most edge weights were estimated accurately.

Figure B2 presents the outcomes of the bootstrap analysis to assess the stability of the centrality indices, that is, node strength, betweenness, and closeness. It depicts the correlation between the bootstrapped mean centrality score and the original network score over increasingly smaller bootstrapped subsamples. As can be seen, the node strength index was much more stable than the other two indices. The low stability of betweenness and closeness is consistent with previous research showing these indices to be unstable in psychological networks (Bringmann et al., 2019).

Figure B2

Subsetting Bootstrap for the Centrality Indices



Note. The dark orange line represents the stability of the betweenness index; the light green line represents the stability of the closeness index; the light blue line represents the stability of the node strength index. The *x*-axis represents the percentage of participants used for the analyses after dropping random participants from the original data set; the *y*-axis represents the average correlation of the centrality order of the subset sample and the centrality order of the original sample. See the online article for the color version of this figure.

Received June 9, 2021
Revision received April 11, 2022
Accepted May 15, 2022 ■