Applied Psycholinguistics (2021), **42**, 527–548
doi:10.1017/S0142716420000533

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Measuring bilingualism: The quest for a "bilingualism quotient"

Viorica Marian* and Sayuri Hayakawa

Northwestern University
*Corresponding author: E-mail: v-marian@northwestern.edu

(Received 31 January 2020; revised 13 July 2020; accepted 28 August 2020; first published online 11 November 2020)

Abstract
The study of bilingualism has a history that extends from deciphering ancient multilingual texts to mapping the structure of the multilingual brain. The language experiences of individual bilinguals are equally diverse and characterized by unique contexts of acquisition and use that can shape not only sociocultural identity but also cognitive and neural function. Perhaps unsurprisingly, this variability in scholarly perspectives and language experiences has given rise to a range of methods for defining bilingualism. The goal of this article is to initiate a conversation about the utility of a more unified approach to how we think about, study, and measure bilingualism. Using concrete case studies, we illustrate the value of enhancing communication and streamlining terminology across researchers with different methodologies within questions, different questions within domains, and different domains within scientific inquiry. We specifically consider the utility and feasibility of a bilingualism quotient (BQ) construct, discuss the idea of a BQ relative to the well-established intelligence quotient, and include recommendations for next steps. We conclude that though the variability in language backgrounds and approaches to defining bilingualism presents significant challenges, concerted efforts to systematize and synthesize research across the field may enable the construction of a valid and generalizable index of multilingual experience.

Keywords: bilingualism; individual differences; language experience; language proficiency

Measuring bilingualism

Introduction

"How bilingual are you?" Anyone who speaks multiple languages is likely to have been questioned about their experiences and abilities—questions such as how, when, and why different languages were acquired. Each of these queries can be deceptively difficult to answer, but attempting to quantify "how bilingual" is an individual (especially relative to others) is particularly challenging due to the many forms that bilingualism can take. Consider, for instance, a native English speaker who began learning Spanish in college, who has little exposure to native Spanish

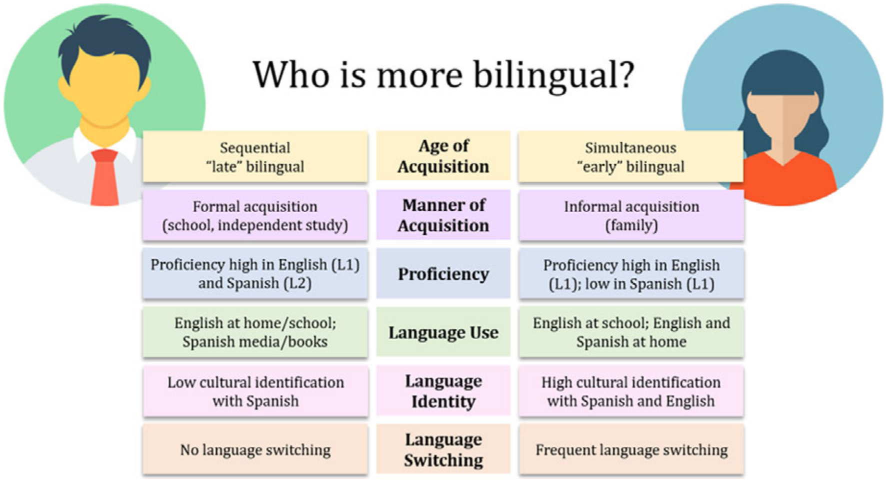

Figure 1. Illustration of the variability in language profiles between two bilinguals.

speakers, but has attained high levels of proficiency reading and writing due to continued independent study and engagement with Spanish literature. Now consider someone exposed to both English and Spanish from birth, who frequently code switches within the home, but has a limited Spanish vocabulary due to the fact that English is the primary language everywhere else (see Figure 1). These two bilinguals differ in their age and manner of acquisition, proficiency, language use, language switching, and language identity—but which of them would be considered "more" bilingual? Though the answer may not be of special importance to any given individual, the ability to quantify and compare levels of bilingualism is critical for those conducting research with bilingual populations. The goal of this article is to pose the question of whether the essence of bilingualism can or should be captured using a single quotient and, if not, how we can work toward gaining a more complete characterization of language background. To this end, we provide an overview of existing methods and challenges for bilingualism researchers, and call for greater communication and consensus as a way to move forward.

Quantifying bilingualism: Self-reports and standardized tests

To address the need for reliable measures of bilingualism, the Language Experience and Proficiency Questionnaire, or LEAP-Q (Marian et al., 2007) was developed and made freely available online for the increasingly linguistically diverse scientific landscape. Ten years later, the authors published a summary describing its use over a decade of research (Kaushanskaya et al., 2019). The LEAP-Q has now been cited more than 1,200 times, used in hundreds of studies worldwide, adapted for use with children, and translated into 26 languages, including Arabic, French, Mandarin, Russian, and Turkish (see Figure 2). This self-report survey, which is available as a paper-and-pencil assessment, as well as electronically, includes measures of

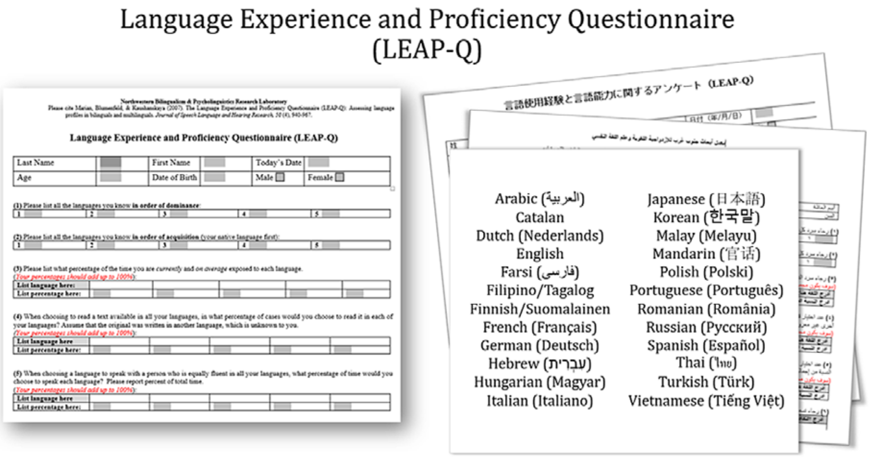

Figure 2. The Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) is available for use with children and adults, can be administered as an online or paper/pencil survey, and has been translated into 26 languages. The LEAP-Q is freely available for download at http://bilingualism.northwestern.edu/leapq/.

language proficiency, age of acquisition (AoA), quantity and type of language exposure, cultural identification, and more.

Other labs around the world have developed similar self-report questionnaires to quantify bilingualism. One commonly used tool is the Language History Questionnaire (Li et al., 2006), which is well suited for online research due to its user-friendly web interface. Another option is the Language Background Questionnaire (Sabourin et al., 2016), which includes a short and long version, allowing researchers to flexibly utilize measures of varying depth and detail depending on the population and topic of interest. The recently updated Language and Social Background Questionnaire (Anderson et al., 2018; Luk & Bialystok, 2013) includes detailed questions regarding language use in specific contexts, such as browsing the Internet or healthcare services, and with specific individuals, such as with neighbors or grandparents. In addition to assessing how often different languages are used in different contexts, efforts have been made to quantify the extent to which individuals switch between different languages within and across contexts. For instance, Gullifer and Titone (2019) devised a method for calculating a continuous measure of "social diversity of language use" (available in the languageEntropy R package) based on what they describe as "language entropy"—that is, how predictably a particular language will be used within a given context. Other self-report questionnaires have been developed to obtain more fine-grained measures of language switching behavior. For instance, the Bilingual Switching Questionnaire (Rodriguez-Fornells et al., 2012) can be used to quantify how often individuals switch into L1 versus L2, how often switches are triggered by particular situations or topics, as well as how often switches occur unintentionally. There are also validated tools specifically designed for use with children that rely on parental reports, such as the Language Exposure Assessment Tool (DeAnda et al., 2016) and the Alberta Language and Development Questionnaire (Paradis et al., 2010), as well as surveys that include both parental and teacher assessments of children's language background and proficiency (e.g., Gutiérrez-Clellen & Kreiter, 2003).

One advantage of self-report measures is the ability to efficiently tap into a number of complex constructs spanning a range of dimensions relevant to bilingual status. However, there are several limitations, including response biases and differences in how respondents interpret questions and scales (Tomoschuk et al., 2019). While self-reported measures of proficiency often show robust correlations with objective measures (Marian et al., 2007; Ross, 1998), the validity of self-reports can vary considerably across skills (e.g., reading vs. writing; Ross, 1998), measures (e.g., specific tasks vs. general abilities; Brantmeier et al., 2012), proficiency levels (Delgado et al., 1999), ages (Bedore et al., 2012), and raters (e.g., parents vs. teachers; Bedore et al., 2011). Therefore, the prevailing recommendation within the field has been to utilize a combination of self-report questionnaires and standardized tests to gain a more accurate and comprehensive picture of bilingual ability and experience. Some of the most frequently utilized standardized tests include picture naming and expressive vocabulary tasks, such as the Expressive Vocabulary Test (Williams, 1997), Multilingual Naming Test (Gollan et al., 2012), and Boston Naming Test (Kaplan et al., 1978), as well as tests of receptive vocabulary, such as the LexTALE (Lemhöfer & Broersma, 2012) and Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). In recent years, researchers have also taken advantage of newer technologies to obtain real-time measures of language behavior, for example by sampling speech throughout the day using tools like Electronically Activated Recorders (Mehl et al., 2001), Language Environment Analysis (Gilkerson & Richards, 2008), and Ecological Momentary Assessment (Jylkkä et al., 2020), which can record language use and prompt participants to report language behavior at specified times using smartphone applications. Despite the increasing availability of useful tools for describing and quantifying bilingualism, a common question that often arises is how researchers can obtain a single score to capture an individual's level of bilingualism.

The quest for a quotient

The quest for a "bilingualism quotient" (BQ) is reminiscent of scientific endeavors to find a quantifiable measure of intelligence. As with bilingualism, which has been, and continues to be, operationalized using highly heterogeneous methods, the study of intelligence has been approached from numerous perspectives, ranging from the early and now-defunct use of craniometry to measure skulls to the standardized cognitive measures seen today (see Figure 3). A critical development in the quantification of intelligence was the introduction of the Binet–Simon intelligence test (Binet–Simon, 1905), which assessed abilities such as logical reasoning and picture naming. These measures were then used to derive a single value representing an individual's "mental age" relative to what would be expected based on one's chronological age and, in contemporary versions, relative to one's age cohort. The later adapted Stanford–Binet test (1916) remains a popular tool and has provided a model for how complex traits and abilities could plausibly be distilled down to a single quotient.

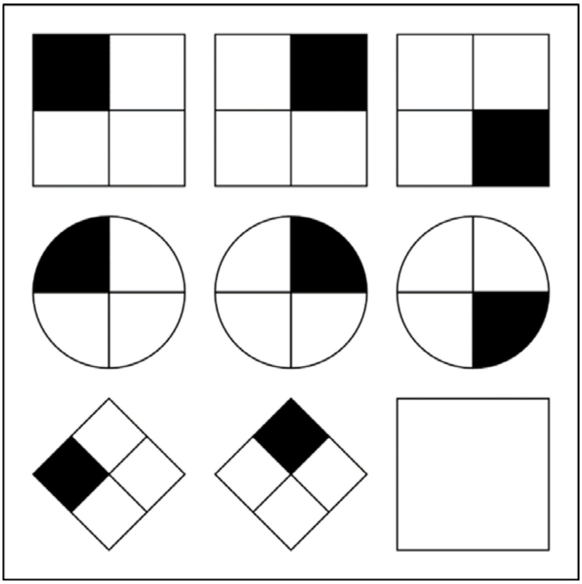

Figure 3. Example item from a cognitive ability test typically used to measure an individual's IQ (based on Raven's Progressive Matrices; Raven, 1936; Wikimedia Commons/CC-BY-SA-3.0).

And yet, Binet has stressed the limitations of such measures and the "difficulty in expressing all the oscillations of thought in a simple, brutal number" (Binet, 1900, p. 119–120).

While IQ tests have undoubtedly served important functions, researchers, educators, and clinicians have increasingly begun to question whether a single number is sufficient to capture the many facets and nuances of intelligence (Sternberg et al., 2001; Stuebing et al., 2002). This has given rise to theories centered on the notion of multiple intelligences, such as Sternberg's (1985) triarchic theory of intelligence, which posits the existence of multiple subcomponents that are broadly grouped under analytical, creative, and practical intelligence, as well as Gardner's (1983) theory of multiple intelligences, which further specifies different domains of intelligence including musical, bodily-kinesthetic, inter- and intrapersonal, linguistic, logical-mathematical, and spatial intelligence. Intelligence is, therefore, a complex construct that has warranted extensive research and inspired countless articles and books—some of which, like *The Bell Curve* (Herrnstein & Murray, 1996), introduce problematic ideas such as the existence of "cognitive elites" and group differences in intelligence. Reducing a multifaceted construct down to a single number can therefore be not only challenging, but even dangerous, as biased assessments can be used and interpreted in nefarious ways. In short, though measures that output a single index of intelligence may have practical value, they are likely to, at best, mask the nuances of human intellectual ability and, at worst, result in unwarranted characterizations of individuals whose intelligence deviates from the models that are imposed upon them.

The search for a BQ presents a similar dilemma, as bilingualism emerges from a complex and interactive set of dimensions that can take many different forms.

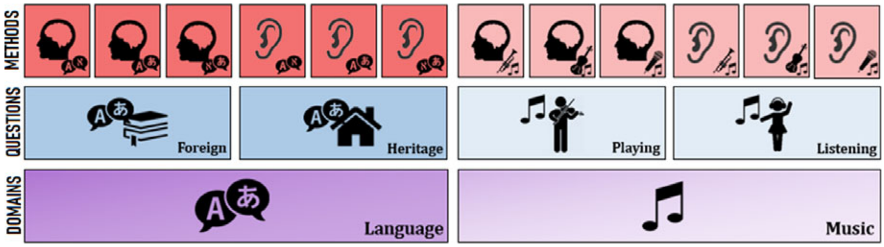

Figure 4. An illustration of the communication hierarchy across methodologies within questions (e.g., effects of bilingualism on executive function), questions within domains (e.g., foreign vs. heritage language learning), and domains within scientific inquiry (e.g., consequences of linguistic experience vs. music experience).

Attempts to quantify an individual as more or less bilingual using a single quotient are therefore meaningless without specifying a particular dimension of interest. Unfortunately, even if we restrict our analyses to a specific variable (such as AoA or proficiency), it is not possible to fully disregard other factors, as the variable of interest is typically related to others, making it difficult to draw causal inferences and generalize across different types of bilinguals. While it may not be reasonable to expect a simple solution to a complicated problem, we propose that enhancing communication and consensus is a critical step toward addressing a number of challenges relevant to quantifying and understanding bilingualism.

We begin by considering the importance of communication across researchers using different methods to study the same question, illustrating it through a discussion of the effects of bilingualism on executive function. Next, we highlight the need for cross talk when asking different questions within the same domain, with foreign versus heritage language learning as the test case. Finally, we discuss the value of greater conceptual integration across different domains of scientific inquiry, using linguistic experience versus musical experience as an example (see Figure 4).

Enhancing communication and consensus

Across methods

A prominent issue within bilingualism research is the heterogeneity of methods and measures used to characterize bilinguals and monolinguals, even among those investigating similar questions. While few researchers are likely to disagree that bilingualism is a continuum rather than a category, it is still common practice to broadly group individuals as being either bilingual or monolingual. In addition to losing potentially valuable information by masking variability within each group, this practice forces researchers to impose an artificial boundary based on criteria that are idiosyncratic across studies (see Bedore et al., 2012; Sabourin et al., 2016 for discussions). Combined with the fact that populations differ dramatically in language history and experience, this heterogeneity in how bilingualism is operationalized can make it difficult to interpret variable findings across different studies. Though such variability can be a significant challenge, it is possible to take advantage of these differences to better understand the consequences of multilingual experience so long as relevant information about participants is measured in ways that allow for comparisons across studies. To foster cross-laboratory communication and consensus, it is necessary to provide detailed descriptions of the populations tested following a consistent approach.

One notable example of problems that arise from lack of standardization can be observed in the so-called bilingual advantage controversy in executive function. While many studies have provided evidence suggesting that multilingual experience can confer advantages for domain-general cognitive functions (e.g., Abutalebi et al., 2011; Blumenfeld & Marian, 2014; see Bialystok, 2009 for a review), these effects are not consistently replicated, leading some to suggest that no such advantages exist (e.g., Dick et al., 2019; Paap & Greenberg, 2013). As noted by many in the field, however, it is difficult to interpret these inconsistent results due to great variability in how both executive function (Hilchey & Klein, 2011; Valian, 2014) and bilingualism (Luk & Bialystok, 2013; Marian, 2018) are defined and measured across studies. Factors extraneous to language experience, such as socioeconomic status (SES) and education, can mask or masquerade as effects of bilingualism when comparing groups that are likely to differ in ways other than bilingual status. Even when these factors are well controlled, relying on a single dimension of bilingualism such as AoA or proficiency to explain variance in executive function can lead to misleading conclusions, as these variables are likely to be correlated with other potentially unmeasured components of language experience (Kaushanskaya & Prior, 2015; see also DeLuca et al., 2019a, 2019b for examples of how different forms of bilingual experience affect neural structures).

Though it may not be possible to account for all relevant variables, it is possible to narrow down the relevance of particular features of bilingualism through the use of comprehensive language assessments and analyses that are more sensitive to interactions among variables. We can use the LEAP-Q to illustrate this. As noted, the LEAP-Q assesses a broad range of variables relevant to language learning and use, including AoA, immigration experience, language exposure, and proficiency in each known language. To validate the measure, self-reported proficiency was compared to scores on objective, standardized tests, revealing a high degree of correspondence. We then utilized factor analysis to identify three orthogonal factors, which were identified as L1 competence (e.g., proficiency and preference for L1), late L2 learning (e.g., age of L2 acquisition and years in an L2 country), and L2 competence (e.g., proficiency and preference for L2). These types of analyses are useful for determining whether measures capture variables of interest, as well as to reveal the underlying associations between variables. It is critical, however, to keep in mind that the relationships between language variables will vary across populations and contexts—factors extracted from a particular population should not be assumed to reflect a universal characterization of bilingualism. Rather, researchers should collect detailed information on their populations and determine the relationship among variables for particular cases, including the relationship between language experience and executive function (see Iverson & Miller, 2017 and Miller & Rothman, 2020 for similar approaches to studying language acquisition, attrition, and maintenance). Other analyses, such as mixed-effects modeling (e.g., De Cat et al. 2018; Linck & Cunnings, 2015), can be used to estimate the contributions of specific aspects of bilingual experience while accounting for variance explained by participant and item-related factors. There are also methods such as propensity matching, whereby pairs of participants are matched across multiple attributes (e.g., SES, musical ability) to reduce the impact of confounding variables (e.g., Hartanto & Yang, 2019). In this respect, longitudinal investigations can be particularly informative, as the effects of bilingualism can be observed over time while controlling for individual differences (Hope, 2015).

These types of approaches can help us make strides toward identifying the aspects of bilingual experience that affect domain-general cognitive abilities, but a number of limitations remain as a result of heterogeneous participant populations and research methods. As noted, relationships among linguistic and nonlinguistic variables may not generalize from one population to another (Adesope et al., 2010). Even the relationship between self-reported and objective measures can vary across populations (Edele et al., 2015). The considerable variance in how different factors relate to each other illustrates that there is no one-size-fits-all method or tool that can be universally applied to all bilinguals. Until there is greater consensus on how individual differences should be assessed, it will be challenging to determine whether inconsistent results across studies emerge from differences in populations or measures. It is, therefore, necessary to develop methods to equate indices derived from different assessments and actively strive to build agreement around basic, but important, procedural and conceptual tenets. In a recent commentary, Leivada et al. (2020) identify procedural characteristics that could help facilitate comparisons across studies investigating the consequences of bilingualism. These include obtaining measures of individual differences and language characteristics, as well as the use of registered reports and appropriate control groups. The authors propose that multilab teams could systematically examine the impact of particular factors by intentionally holding key variables, such as the target language group, constant across research teams.

Though the precise constellation of relevant variables will continue to vary across participants and researchers, establishing consensus around terminology and best practices, including the types of information that should be reported in publications, will facilitate cross-laboratory comparisons and collaborations. It is important to note, however, that we are not advocating for a single static set of variables to be examined and reported. Such a rigid approach could contribute to reductionist characterizations of bilingual experience and impede scientific discovery (e.g., of new relevant variables). Rather, a potential solution could be to cultivate norms around a *minimum* set of attributes that should be reported to facilitate comparisons across studies, akin to variables such as age and gender, which are commonly reported regardless of the research question. By enhancing communication and consensus across laboratories investigating similar phenomena, we can move away from piecemeal investigations to gain a broader, more complete understanding of the consequences of multilingual experience.

Across questions

In addition to establishing consensus when studying the same phenomenon, our understanding of bilingualism can benefit from enhancing cross talk between researchers asking different questions. As it is, even seemingly similar spheres of investigation often suffer from lack of integration across their respective literatures (Kondo-Brown, 2005). There is often limited consideration of work outside the narrow scope of a research agenda, even in cases in which the same types of tasks are utilized. As a result, we potentially miss out on opportunities to learn from relevant, but distinct, programs of research that may have already generated a wealth of information that could help us develop more sensitive, reliable, and comprehensive measures of individual differences. Researchers within areas of study come to acquire particular biases and ways of thinking about a question. We can capitalize on cross-area variability in approaches and perspectives to reveal blind spots within our own areas, as well as identify common mechanisms and potential moderators.

Take, for instance, research examining foreign versus heritage language learning. In both areas, researchers often seek to identify individual differences in language background that predict language proficiency and maintenance. However, in addition to limited standardization within each subfield (Norris & Ortega, 2000; Porte, 2012), the two areas often differ in the types of variables that are examined. While foreign language studies commonly look at factors such as AoA, language exposure, and domain-general cognitive abilities (e.g., Birdsong, 2006; DeKeyser, 2005), research with heritage learners is far more likely to consider the impact of sociocultural identity, as well as personal and community attitudes toward particular languages (Oh & Au, 2005; Potowski, 2004). Given that these variables can impact language learning, it would be valuable to consider their potential influence in both subfields. For instance, language identity has been associated with motivation and investment in a language for both heritage and foreign language learners, which in turn can promote language acquisition and maintenance (Norton, 2000; Potowski, 2004). Similarly, factors such as the quality and quantity of formal training can have a significant impact on language abilities, not only among foreign language learners but for heritage speakers as well (Kupisch & Rothman, 2018; Putnam et al., 2018).

It is equally appropriate to determine the validity of adopting measures designed for different types of populations. It is not uncommon for heritage language researchers to utilize proficiency tests designed for foreign language learners, which could potentially result in misleading characterizations of language ability (Montrul & Perpinan, 2011; Potowski et al., 2012). For instance, Montrul and Perpinan (2011) suggest that written production tests designed for foreign language learners may underestimate grammatical ability among heritage learners because the latter often relies more on implicit knowledge, which may not be captured by this type of task. The issue of utilizing possibly inappropriate diagnostic tools is reminiscent of problems that emerge when language assessments designed for monolingual populations are adopted for bilingual learners, leading to potentially inaccurate characterizations of language differences as language disorders (Gutiérrez-Clellen & Peña, 2001). Even seemingly straightforward variables can be difficult to operationalize when dealing with heterogeneous populations (de Bruin, 2019). For instance, as noted by Unsworth and Blom (2010), age of first exposure may be relatively simple to determine when there is a clear-cut age of immigration, but there are often complicating factors, such as if an individual's immersion in a given language community is interrupted. It may therefore be necessary to acquire different or additional information depending on the population. Determining the appropriateness of particular tests for different populations will not only allow researchers to develop better assessments but also facilitate comparisons of different language groups. This could have important practical implications, such as for language education programs (e.g., Kondo-Brown, 2003), as well as theoretical value, such as for investigating critical periods (e.g., Benmamoun et al., 2013; Judy et al., 2018).

While some progress has been made to identify commonalities and differences across foreign and heritage language learners, more research is needed to systematically evaluate how reliably assessments designed for particular populations can be generalized to other groups (Hulstijn, 2012), as well as to develop guidelines for how to adapt and validate questionnaires for new populations (Beaton et al., 2000). There is also a need for greater cross talk between researchers, educators, and clinicians so that professionals in applied domains can benefit from empirical research, and research questions and methods can be informed by bilinguals in the real world. In sum, greater integration across areas of bilingualism research can not only expand our current understanding of bilingualism but also facilitate the development of more comprehensive and valid tools as we all move forward.

Across domains

So far, we have considered how greater communication across laboratories and areas of bilingualism research could improve measures of language background and clarify the consequences of bilingual experience. Isolating the impact of language experience, however, requires consideration of how individuals differ in ways other than language background. Different forms of experience—such as music, physical exercise, and video-game playing—can have an impact on both linguistic and nonlinguistic abilities, which, if not accounted for, can be confounded with the effects of bilingualism. Effects of language and other forms of experience can also emerge from shared mechanisms, which could facilitate the development of theoretical frameworks and justifications for examining specific variables of interest.

Take, for instance, the often-overlapping consequences of language and musical experience for executive function and working memory (Janus et al., 2016; Parbery-Clark et al., 2009), as well as auditory processing and perception (Krizman et al., 2014; Ressel et al., 2012). Language and musical ability are not necessarily independent. There is evidence that processing music engages neural regions commonly associated with the language network (Maess et al., 2001), and ERP components reveal similar processing of musical and linguistic syntactic anomalies (Patel, 2003). Simultaneous presentation of incongruent musical and linguistic pitch contours can modulate the N2b component among speakers of tonal languages (Sadakata et al., 2020), suggesting that listening to music has the potential to actively interfere with concurrent language processing. Longer-term musical training has also been found to influence language learning (Zhao & Kuhl, 2016), often resulting in beneficial effects, such as for speech perception. Language experience can influence musical ability, for instance by increasing sensitivity to pitch contrasts (Liu & Kager, 2017).

Compared to other forms of experience, however, the relationship between language and musical abilities has become a relatively fertile area of research and, on several occasions, researchers have directly compared individuals with bilingual

Table 1. Mean interference effects (incongruent–neutral RT) and participant demographics across groups in a study examining the effects of bilingual experience and music experience on cognitive control

	Bilinguals (N = 43)	Musicians (N = 42)	Bilingual Musicians (N = 69)	Control (N = 65)
Interference (ms; Simon)*	14.47 (2.58)	17.16 (2.66)	16.71 (2.34)	28.68 (2.58)
Age (years)*	22.30 (4.05)	22.21 (3.42)	20.60 (2.86)	22.88 (4.03)
IQ (WASI)*	109.16 (8.59)	114.74 (9.92)	113.55 (9.56)	111.67 (11.39)
Digit Span (CTOPP)*	16.12 (3.00)	17.69 (1.54)	17.47 (2.76)	17.55 (2.46)

Note: Values in parentheses represent standard errors of mean differences for interference and standard deviations for age, IQ, and digit span. Asterisks indicate significant group differences at $p < .05$. Adapted from Schroeder et al., 2016.

and/or musical experience. To illustrate, Schroeder et al. (2016) investigated the independent and additive effects of bilingual and musical experience on executive control using the nonlinguistic, nonmusical Simon task. Critically, they compared four groups of individuals—bilinguals, musicians, bilingual musicians, and nonbilingual nonmusician controls (see Table 1). The authors observed that both bilingualism and musicianship improved cognitive control, but that having both types of experience did not confer any additional advantages. These types of comparisons across domains can help inform us about the potential mechanisms underlying the respective effects and help dissociate their relative contributions. There have also been longitudinal studies employing language and music training paradigms (e.g., Bak et al., 2016; Janus et al., 2016), which are valuable for determining causal relationships.

In many cases, however, it remains difficult to draw causal inferences as it is not always feasible to directly manipulate musical and language experience (see Schellenberg, 2019). It can also be difficult to equate levels of language versus musical experience, which can diminish the ability to interpret cases in which effects are found for one and not the other. There has additionally been relatively little consideration of other sources of individual variability, such as physical exercise or video-game playing (Valian, 2015), both of which can have similar consequences to those of music and language (Bavelier et al., 2012; Colcombe & Kramer, 2003). All this to say that bilingualism researchers would benefit from greater consideration of measures assessing other sources of individual differences, both to control confounding variables and to provide insight into underlying mechanisms—systematically examining how and when different forms of experience have common or distinct consequences for cognitive and neural functioning. These insights can, in turn, enable the development of more precise theoretical frameworks of bilingualism and cognition.

A bilingualism quotient?

Greater communication and consensus, both within and across disciplines, can help us converge upon the processes underlying the impact of bilingualism and the power of experience more generally. Identifying how different components of bilingualism relate to each other, as well as how they systematically vary across groups and contexts, could also help us develop tools that are suitable for a wider range of purposes—potentially even culminating in the elusive BQ.

What do IQ and BQ measure?

We began by drawing the parallel to the field of intelligence research and, in particular, the development of IQ as a seemingly tenable model for how one could distill complex constructs down to a single number. The concept of IQ is predicated on the idea that there is a general mental capacity—often identified as the G factor—that influences other, more specific cognitive abilities (Spearman, 1904). IQ is thus derived by extracting the shared variance among a battery of tests (e.g., verbal fluency, mental rotation), with the logic that the latent factor common to all of them indexes general intelligence. Though bilingualism is similarly associated with multiple, often correlated attributes, many of them characterize language history or context, rather than language ability per se (see Figure 5 for an example). This is in contrast to the subcomponents of intelligence, which uniformly pertain to cognitive functions that could reasonably be expected to emerge from a common underlying capacity, such as G. In other words, while the G factor is thought to be the *cause* of associations among different components of intelligence, a B factor is more likely to be an *emergent property* of existing relationships (e.g., between AoA and proficiency). This distinction is meaningful, as an effect of BQ could tell us that some aspect of bilingualism is relevant to a variable of interest, but not necessarily what that aspect is.

Both intelligence and bilingualism have also been notoriously difficult to operationalize due to their multidimensional nature and associated outcomes. Historically, intelligence has been considered an *ability*, and yet IQ is highly correlated with standardized tests of *achievement* (Frey & Detterman, 2004) and academic success (Deary et al., 2007). Bilingualism is often considered an *achievement*, but impacts a host of *abilities*, ranging from speech discrimination (Garcia et al., 2018) to language learning (Kaushanskaya & Marian, 2009). Both constructs also share reciprocal relationships to environmental variables such as education and SES (Neisser et al., 1996; Peng et al., 2019; Rindermann et al., 2010; Wyche et al., 2019)—further blurring the lines between assessing ability and achievement and raising fundamental questions about what, exactly, intelligence and bilingualism measure.

How would BQ be computed?

Beyond determining what a BQ is meant to capture, a notable challenge will be to decide what it means to be more or less bilingual. Variability in some, if not all, relevant attributes (e.g., manner of acquisition and use) are more naturally conceived of as *qualitative* differences that do not necessarily index "degree" of bilingualism. Though it is possible to operationalize most variables along a continuum (e.g., more or less immersion), calculating a single quotient almost necessarily requires a value judgment. One approach could be to make such determinations on a semantic basis, such as by classifying greater frequency, duration, and depth

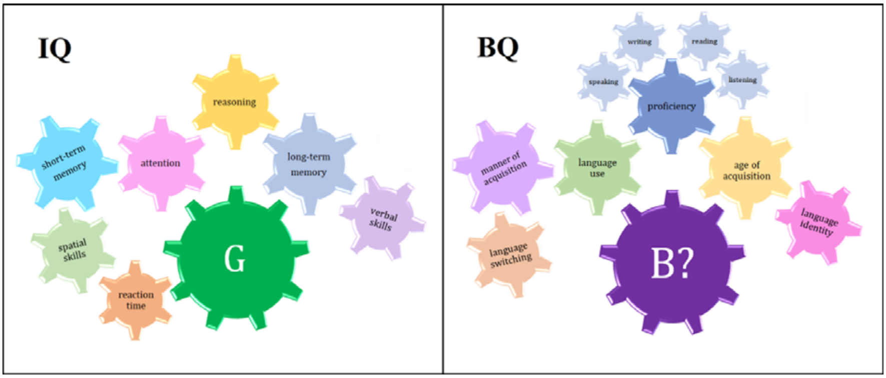

Figure 5. Schematic of relations among sample variables relevant to intelligence (left) and bilingualism (right).

of bilingual experience as "more" bilingual. We could also empirically identify the characteristics that explain variability in other domains, such as neural function. These classifications could then serve as the basis for a BQ, which could in turn be used to explore other consequences and antecedents of bilingualism. The danger of this method, however, would be in the potentially tautological use of a BQ to investigate its relationship to the very functions or variables used to construct it.

A potentially simpler option may be to restrict BQ to cognitive abilities, which are easier to conceive along a continuum. Using large datasets, we may be able to identify a general index of language ability that supports specific functions (e.g., oral production, literacy). Such a metric could additionally inform the results of earlier factor analyses (Marian et al., 2007) indicating that different facets of language proficiency (e.g., speaking, writing, reading) are associated with different aspects of language history (e.g., AoA, L2 exposure, L1 duration, respectively; see Figure 6). One of the main challenges will be to determine whether there are, in fact, metrics that have strong predictive validity across domains and populations—and if not, whether there are systematic moderators that can be used to "customize" the formula.

Developing a single, generalizable index of bilingualism could make a significant contribution to the field by facilitating direct comparisons across individuals, as well as across laboratories. Such a tool does, however, have the limitation of obscuring the many cognitive, affective, social, and environmental dimensions of bilingualism that do not directly characterize language ability. Assuming it is possible to use a more diverse set of variables as metrics of bilingualism, the BQ would ideally maximize flexibility over the choice of particular measures, while still maintaining an overall level of consistency.

For instance, we could start by identifying four to six general categories corresponding to key components of bilingual experience. These categories would be consistent and agreed upon by researchers in the field and may include, for example, language ability, history of use, acquisition, and possibly aspects like identity or cultural affiliation or others. Each of these categories would include multiple measures within it. The language ability category, for example, could include abilities to

Factor 1: L1 Competence	Loading values	Factor 2: Late L2 Learning	Loading values	Factor 3: L2 Competence	Loading values	Factor 4: L1 Maintenance	Loading values
Proficiency reading	.947	Age became fluent	.864	Exposure (% time)	.923	L1 exposure to classes	.916
Comfort understanding	.910	Age began acquiring	.859	Exposure to TV	.908	L1 exposure to TV	.914
Proficiency understanding	.910	Age became fluent reader	.855	Exposure to friends	.861	L1 exposure to radio	.831
Comfort writing	.903	Comfort understanding	−.803	Exposure to radio	.772	L1 exposure to reading	.776
Proficiency writing	.896	Age began reading	.751	Writing proficiency	.660	L2 learning from reading	−.727
Comfort reading	.884	Proficiency understanding	−.697	Exposure to family	.621	L1 exposure (% time)	.627
Identified accent	−.788	Years in a country	−.681	Comfort writing	.592	L1 exposure to friends	.530
Comfort speaking	.748	Learning from tapes	.601	Preference to speak	.590		
Proficiency speaking	.704	Proficiency speaking	−.580	Exposure to reading	.564		
Cultural identification	.526			Exposure to classes	.543		
Perceived accent	−.517			Learning from reading	−.519		
Preference to read	.457						
% variance	23.480		13.383		9.625		7.534
Cumulative variance	23.480		36.862		46.488		54.021
Factor 5: Late L2 Immersion	**Loading values**	**Factor 6: Media-Based Learning**	**Loading values**	**Factor 7: Non-Native Status**	**Loading values**	**Factor 8: Balanced Immersion**	**Loading values**
L1 years of class learning	.728	L1 learning from TV	.866	L2 perceived accent	.839	L1 learning from friends	−.813
L2 years in workplace	.725	L2 learning from TV	.838	L2 identified accent	.615	L2 years of schooling	.627
L1 years in workplace	.714	L1 learning from the radio	.741	L2 cultural identification	−.602	L1 years in a family	.622
Proficiency reading L2	−.687	L2 learning from the radio	.652	L1 age became fluent	.590	L2 years in a classroom	.541
L2 learning from friends	−.683	L2 comfort reading	−.476	L2 learning from family	−.519	L1 years in a country	.499
L2 learning in a classroom	−.556						
L1 years in school	.476						
% variance	6.424		6.226		5.049		4.232
Cumulative variance	60.445		66.671		71.720		75.952

Figure 6. Results of a factor analysis on measures of the LEAP-Q. Reproduced from Marian et al. (2007; Table 2).

comprehend spoken language, to speak, to read, to write, and perhaps other abilities as well, as agreed upon in the field. Individual researchers could then have the freedom to select specific submeasures from each category that are most relevant for their research question and population under study. For example, from the acquisition category, for some research questions *age of acquisition* may be the most relevant measure, whereas for other research questions the *manner of acquisition* (e.g., classroom or heritage) may be more relevant. Similarly, from the history of use category, for some research questions *current use* may be most relevant, whereas for other research questions *cumulative historical use* may be more relevant. The exact number of measures (one, two, more) from each category that is to be used toward computing the BQ would be determined during the course of BQ development. Responses could then be weighted and aggregated to obtain a single quotient. The BQ will thus be a composite score across the agreed-upon general categories, where the score for each general category is obtained based on the subdimensions from that category that are most relevant to the specific researcher. This will provide the advantage of flexibility for individual researchers to choose the variables most relevant for their topic from each category while maintaining uniformity by indexing all agreed-upon general categories within the BQ construct.

The advantage of including flexible options would be that the same categories could be used for a wide range of purposes and, hopefully, result in widespread adoption of a common construct. Constraining the represented categories, however, would help facilitate cross-laboratory comparisons.

Variability across time and groups

In the final part of this section, we consider how we might increase the validity of a BQ by addressing two complicating factors—variability across time and variability across groups.

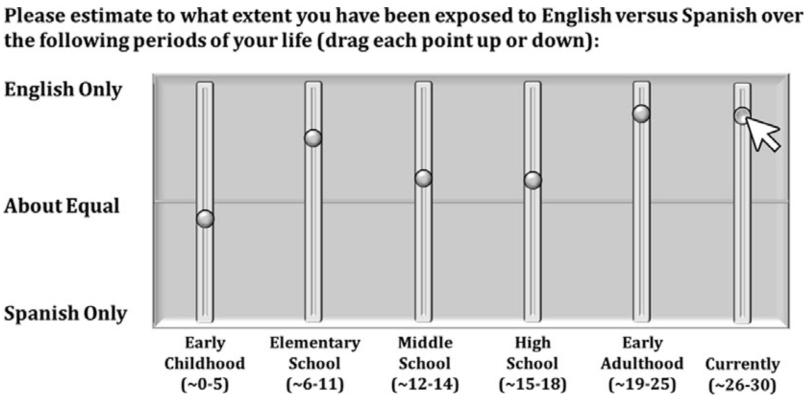

Figure 7. Example of a developmental self-report item assessing relative language exposure over time for a 30-year-old English–Spanish bilingual.

Intelligence research shows that IQ is relatively stable over one's lifetime, and the genetic versus environmental contributions to IQ have been studied extensively. In fact, it is this stability over one's lifetime that made IQ scores so attractive historically—to the military, potential employers, and academic institutions. Such stability cannot be expected, and in fact is directly inconsistent with the dynamic nature of bilingualism. Even cognitive abilities such as language proficiency are expected to fluctuate over time as a bilingual learns or forgets a language (e.g., due to immigration, international adoption, or other changes in language use). Though it may not be possible to predict future circumstances, test items could be designed to assess levels of language proficiency, exposure, use, and so forth at different periods in the past (see Figure 7 for an example). Responses could then be aggregated to calculate a single index, used to classify individuals based on their developmental trajectory, or analyzed individually to determine the impact of particular variables at specific time points.

Extending IQ's cohort-based approach, a BQ could characterize an individual's level of bilingualism relative to others who are matched on key demographic variables such as socioeconomic status and age. Unlike IQ, however, which is generally calculated based on a static set of criteria, a BQ would ideally reflect the dynamic nature of bilingualism through a more flexible system of measurement and scoring. For instance, factors like degree of language exposure and proficiency could be differentially weighted depending on demographic and language background variables. Using age as an example, assessments of young children would likely privilege factors such as language exposure over language proficiency, while measures of linguistic competence may be weighted more heavily for older children (see Byers-Heinlein et al., 2019). We could rely on the data to determine the weights for a given study. Factor loadings could be determined by the intercorrelations among variables for a particular population, so that the variables that share the most variance with all others would be weighted the most. This approach is similar to the one behind IQ, such that the BQ would represent the latent aspect of bilingualism that is common across factors. In addition to allowing for "customized" factor loadings, this

Table 2. Summary of key challenges, points of discussion, and recommendations

Challenges
• BQ would need to account for heterogenous variables related to language ability, history, and context. • Language exposure, use, and ability change over time, and the relationships among variables vary across populations. • Some language background variables capture qualitative differences that do not clearly index a relative “degree” of bilingualism.
Discussion Questions
• What would a BQ represent? • Is it possible to operationalize all relevant factors along a continuum of bilingualism? • What criteria and principles could be used to establish a bilingual continuum?
Recommendations
• Use of large, diverse datasets and advanced statistical methods to select and weight attributes. • Implement flexible methods of administering, weighting, and scoring (e.g., rely on several agreed-upon categories with multiple specific attributes within each to select among). • Variable standards of comparison based on demographic characteristics to determine relative bilingualism. • Include mechanisms for assessing change over time. • Establish norms and procedures for reporting and depositing data.

approach should still enable comparisons across studies, as each participant’s score would represent a relative ranking within their group based on attributes that were most important for that population. It may even be possible to additionally include ways to calculate a score in relation to an established standard.

Conclusions: Toward a unified approach to measuring bilingualism

In summary, we have outlined some potential approaches, challenges, and points of discussion for formulating a tool that could be used to calculate a BQ (see Table 2). We caution, however, that a BQ, while providing a useful heuristic, is subject to the same constraints and limitations as its cousin the IQ on which it was modeled. Just like a single IQ number cannot capture the nuanced cognitive capacities of every mind, especially for neurodiverse populations, a single BQ number will only provide an approximate estimation of a bilingual’s linguistic profile and will have limited predictive ability for specific language acts and individuals. Most proximally, it will be necessary to generate agreement around the form and function of the quotient, possibly through flexible methods of administration and scoring while enabling cross-laboratory comparisons. Advocating for how (or whether) a BQ should be formulated, however, would be antithetical to our ultimate intention, which is to *begin a conversation* (rather than provide definitive answers) so that solutions can be found through scientific discourse and research in the field.

In recent years, there has been significant progress in how society and individuals view bilingualism—these changes in attitudes and perceptions are in no small part due to the findings that have been brought to light by researchers who study bilingualism. How we understand bilingualism has also continued to evolve—for instance, a theme that has increasingly emerged is the importance of considering bilingualism as a continuum rather than a category (see Luk & Bialystok, 2013). Operationalizing bilingualism as a category can mask important variability within each group and impede comparisons across studies, as the boundaries between "monolingual" and "bilingual" are often idiosyncratically and arbitrarily drawn. Even in cases when categorization is necessary, it will be important to rely on standards and guidelines that have been agreed upon by researchers within the field. One silver lining that has emerged from the controversies surrounding bilingualism is that they have inspired researchers to reconsider the types of questions that we should be asking—not merely whether bilinguals differ from monolinguals but rather, how, why, and when language experience alters aspects of cognitive and neural function (see Baum & Titone, 2014; Tabori et al., 2018).

While the inherent variability of language experience can pose significant challenges, it can also provide useful insights so long as we are able to organize the information. Enhancing communication and consensus is an important step toward this goal, but how do we do it? Just as we can look to other fields to shed light on potential obstacles, we may find useful solutions by examining how other spheres of scientific investigation have resolved similar problems. For instance, scientists in many fields, including chemistry, biology, health, and food science, often have professional organizations, such as the US Food and Drug Administration (USDA) and the Clinical and Laboratory Standards Institute (CLSI), which help establish guidelines and build consensus. The CLSI offers online workshops, webinars, and certificate programs for the purposes of enhancing the quality and standards of healthcare-related research. Their website also provides information regarding their consensus process, which begins with the identification of problems and projects by experts and community members, followed by the recruitment of panels and councils to develop plans, disseminate relevant information, and obtain comments and votes on preliminary drafts, and ultimately results in the publication of research standards and guidelines (clsi.org). Groups to develop standards have been established in areas ranging from consumer product development to microbiology research. So, might this be possible for bilingualism?

In fact, there are already initiatives in place, such as Bilingualism Matters (bilingualism-matters.ppls.ed.ac.uk), that have the express purpose of bridging bilingualism researchers with each other and members of the public. This rapidly expanding community of researchers has established branches all over the world. At its inception, Bilingualism Matters was primarily established to help connect bilingualism researchers with community members to enable parents and educators to make better, more informed evidence-based decisions about bilingualism and language learning. As it continues to expand and evolve, Bilingualism Matters and other organizations like it may be able to connect bilingualism researchers with each other and provide a centralized structure through which we can begin to establish initiatives for enhancing communication and consensus.

An increasing number of such collaborative endeavors have already begun to emerge. The Quantifying Bilingual Experience (Q-BEx) project is currently underway to develop measures of cumulative and current language experience, with the objective of enabling cross-laboratory comparisons (see https://q-bex.org/). In collaboration with an international team of researchers, speech-language pathologists, and educators, project lead Cécile De Cat aims to create a user-friendly online tool, which will be available in 13 different languages and with varying levels of detail and length. Along similar lines, the Psycholinguistics of Language Representation (PoLaR) lab has initiated projects to identify variables that predict language performance among heritage speakers, which could help inform the development of novel measures (see https://site.uit.no/polar/funded-projects/).

Another critical piece of the puzzle is the establishment of websites and applications to consolidate and disseminate information, including surveys, assessment tools, data, and analysis scripts. A number of useful resources are already available, such as the Open Science Framework (osf.io) and GitHub (github.com) for uploading data and materials; PsychFileDrawer (psychfiledrawer.org) for reporting published and unpublished findings; and university- and laboratory-run websites, such as by the National Heritage Language Resource Center (nhlrc.ucla.edu), for consolidating assessment tools and making them easily accessible. In addition to continuing to encourage the use and development of these types of platforms, a key objective moving forward should be to establish best practices for what and where information should be reported and deposited, with the ultimate goal of having a centralized and user-friendly repository of tools, procedures, and data. Concerted efforts to build on existing structures for the purposes of improving, standardizing, and consolidating measures could help us move toward enhancing consistency in how we talk about, study, and measure bilingualism.

In other words, to gain a unified and comprehensive understanding of multilingualism, we need to start speaking the same language.

Acknowledgments. Preparation of this article was supported in part by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number R01HD059858 to Viorica Marian. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank Drs. Fatih Bayram, Jorge Gonzalez, Jason Rothman, and Marit Westergaard for the invitation to present an earlier version of this article at the September 2019 Capturing and Quantifying Individual Differences in Bilingualism Symposium in Tromso, Norway, as well as symposium participants for insightful discussions on this topic.

References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., Cappa, S. F., & Costa, A. (2011). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, **22**(9), 2076–2086.

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, **80**(2), 207–245.

Anderson, J. A., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, **50**(1), 250–263.

Bak, T. H., Long, M. R., Vega-Mendoza, M., & Sorace, A. (2016). Novelty, challenge, and practice: The impact of intensive language learning on attentional functions. *PloS One*, **11**(4), e0153485.

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, **35**, 391–416.

Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, **35**(5), 857–894.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, **25**(24), 3186–3191.

Bedore, L. M., Peña, E. D., Joyner, D., & Macken, C. (2011). Parent and teacher rating of bilingual language proficiency and language development concerns. *International Journal of Bilingual Education and Bilingualism*, **14**(5), 489–511.

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., Bohman, T. M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition*, **15**(3), 616–629.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, **39**, 129–181.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, **12**(1), 3–11.

Binet, A. (1900). *La Suggestibilité*. Paris: Schleicher Freres.

Binet, A. (1905). *The development of intelligence in children (the Binet–Simon test)*. Trans. Elizabeth Kite. Baltimore: Williams and Wilkins.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, **56**, 9–49.

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition*, **17**(3), 610–629.

Brantmeier, C., Vanderplank, R., & Strube, M. (2012). What about me? Individual self-assessment by skill and level of language instruction. *System*, **40**(1), 144–160.

Byers-Heinlein, K., Esposito, E. G., Winsler, A., Marian, V., Castro, D. C., & Luk, G. (2019). The case for measuring and reporting bilingualism in developmental research. *Collabra: Psychology*, **5**(1), 37.

Colcombe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: A meta-analytic study. *Psychological Science*, **14**(2), 125–130.

DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The language exposure assessment tool: Quantifying language exposure in infants and children. *Journal of Speech, Language, and Hearing Research*, **59**(6), 1346–1356.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, **35**(1), 13–21.

de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, **9**(3), 33.

De Cat, C., Gusnanto, A. & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, **40**(1), 119–151. https://doi.org/10.1017/S0272263116000486

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, **55**(S1), 1–25.

Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B. (1999). Self-assessment of linguistic skills by bilingual Hispanics. *Hispanic Journal of Behavioral Sciences*, **21**(1), 31–46.

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

DeLuca, V., Rothman, J., & Pliatsikas, C. (2019b). Linguistic immersion and structural effects on the bilingual brain: A longitudinal study. *Bilingualism: Language and Cognition*, **22**(5), 1160–1175.

Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., Riedel, M. C., Laird, A. R., & Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nature Human Behaviour*, **3**, 692–701.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, **52**, 99–123.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science*, **15**(6), 373–378.

Garcia, L. N., Guerrero-Mosquera, C., Colomer, M., & Sebastian-Galles, N. (2018). Evoked and oscillatory EEG activity differentiates language discrimination in young monolingual and bilingual infants. *Scientific Reports*, **8**(1), 2770.

Gardner, H. (1983). *Frames of mind.* New York: Basic Books.

Gilkerson, J., & Richards, J. A. (2008). The LENA Developmental Snapshot (Report No. LTR-07-2). Retrieved from LENA Foundation website: http://www.lenafoundation.org/TechReport.aspx/Snapshot/LTR-07-2

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*, **15**(3), 594–615.

Gullifer, J. W., & Titone, D. (2019). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, **94**(1), 1–18.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, **24**(2), 267–288.

Gutiérrez-Clellen, V. F., & Peña, E. (2001). Dynamic assessment of diverse children. *Language, Speech, and Hearing Services in Schools*, **32**, 212–224.

Hartanto, A., & Yang, H. (2019). Does early active bilingualism enhance inhibitory control and monitoring? A propensity-matching analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **45**(2), 360–378.

Herrnstein, R. J., & Murray, C. A. (1996). *The Bell Curve: Intelligence and Class Structure in American Life.* New York: Simon & Schuster.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, **18**(4), 625–658.

Hope, T. M. (2015). The bilingual cognitive advantage: no smoke without fire. *AIMS Neuroscience*, **2**(2), 58–65.

Hulstijn, J. H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, **15**(2), 422–433.

Iverson, M., & Miller, D. (2017). Language attrition and maintenance: Two sides of the same coin? *Linguistic Approaches to Bilingualism*, **7**(6), 704–708.

Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology*, **144**, 84–97.

Judy, T., Putnam, M. T., & Rothman, J. (2018). When bilingualism is the common factor: Switch reference at the junction of competence and performance in both second language and heritage language performance. *Journal of Language Contact*, **11**(3), 590–616.

Jylkkä, J., Soveri, A., Laine, M., & Lehtonen, M. (2020). Assessing bilingual language switching behavior with Ecological Momentary Assessment. *Bilingualism: Language and Cognition*, **23**(2), 309–322.

Kaplan, E., Goodglass, H., & Weintraub, S. (1978). *Boston Naming Test.* Philadelphia: Lea & Febiger.

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2019). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, 1–6.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, **16**(4), 705–710.

Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition*, **18**(1), 27–28.

Kondo-Brown, K. (2003). Heritage language instruction for post-secondary students from immigrant backgrounds. *Heritage Language Journal*, **1**(1), 1–25.

Kondo-Brown, K. (2005). Differences in language skills: Heritage language learner subgroups and foreign language learners. *The Modern Language Journal*, **89**(4), 563–581.

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and Language*, **128**(1), 34–40.

Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, **22**(5), 564–582.

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*, 1–14.

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.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, **38**(2), 202–210.

Linck, J., & Cunnings, I. (2015) The utility and application of mixed-effects models in second language research. *Language Learning*, **65**(S1), 185–207.

Liu, L., & Kager, R. (2017). Perception of tones by bilingual infants learning non-tone languages. *Bilingualism: Language and Cognition*, **20**(3), 561–575.

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.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, **4**(5), 540–545.

Marian, V. (2018). Bilingual research methods. In J. Altarriba & R. R. Heredia (Eds.), *An introduction to bilingualism* (pp. 12–36). New York, NY: Routledge.

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.

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, **33**(4), 517–523.

Miller, D., & Rothman, J. (2020). You win some, you lose some: Comprehension and event-related potential evidence for L1 attrition. *Bilingualism: Language and Cognition*, **23**(4), 869–883.

Montrul, S., & Perpiñán, S. (2011). Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. *Heritage Language Journal*, **8**(1), 90–133.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, **51**(2), 77–101.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, **50**(3), 417–528.

Norton, B. (2000). *Identity and language learning: Gender, ethnicity and educational change*. Essex, UK: Pearson Education Limited.

Oh, J. S., & Au, T. K. F. (2005). Learning Spanish as a heritage language: The role of sociocultural background variables. *Language, Culture and Curriculum*, **18**(3), 229–241.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, **66**(2), 232–258.

Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, **43**(6), 474–497.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience*, **29**(45), 14100–14107.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, **6**(7), 674–681.

Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin*, **145**(2), 189–236.

Porte, G. (Ed.). (2012). *Replication research in applied linguistics*. Cambridge: Cambridge University Press.

Potowski, K. (2004). Student Spanish use and investment in a dual immersion classroom: Implications for second language acquisition and heritage language maintenance. *The Modern Language Journal*, **88**(1), 75–101.

Potowski, K., Parada, M., & Morgan-Short, K. (2012). Developing an online placement exam for Spanish heritage speakers and L2 students. *Heritage Language Journal*, **9**(1), 51–76.

Putnam, M. T., Kupisch, T., & y Cabo, D. P. (2018). Different situations, similar outcomes. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (Vol. **54**, pp. 251–280). Amsterdam: John Benjamins Publishing Company.

Raven, J. C. (1936). *Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive* (MSc Thesis, University of London).

Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Ávila, C., & Sebastián-Gallés, N. (2012). An effect of bilingualism on the auditory cortex. *Journal of Neuroscience*, **32**(47), 16597–16601.

Rindermann, H., Flores-Mendoza, C., & Mansur-Alves, M. (2010). Reciprocal effects between fluid and crystallized intelligence and their dependence on parents' socioeconomic status and education. *Learning and Individual Differences*, **20**(5), 544–548.

Rodriguez-Fornells, A., Kramer, U., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, **2**, 388.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, **15**(1), 1–20.

Sabourin, L., Leclerc, J. C., Lapierre, M., Burkholder, M., & Brien, C. (2016). The language background questionnaire in L2 research: Teasing apart the variables. In Annual Meeting of the Canadian Linguistics Association, Calgary, Canada.

Sadakata, M., Weidema, J. L., & Honing, H. (2020). Parallel pitch processing in speech and melody: A study of the interference of musical melody on lexical pitch perception in speakers of Mandarin. *PLOS One*, **15**(3), e0229109.

Schellenberg, E. G. (2019). Correlation = causation? Music training, psychology, and neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000263

Schroeder, S. R., Marian, V., Shook, A., & Bartolotti, J. (2016). Bilingualism and musicianship enhance cognitive control. *Neural Plasticity*, **2016**, e4058620.

Spearman, C. (1904). General intelligence: objectively determined and measured. *American Journal of Psychology*, **15**(2), 201–293.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, **47**, 1–41.

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal*, **39**(2), 469–518.

Tabori, A. A. T., Mech, E. N., & Atagi, N. (2018). Exploiting language variation to better understand the cognitive consequences of bilingualism. *Frontiers in Psychology*, **9**, 1686.

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, **22**(3), 516–536.

Unsworth, S., & Blom, E. (2010). Comparing L1 children, L2 children and L2 adults. *Experimental Methods in Language Acquisition Research*, **27**, 201–222.

Valian, V. (2014). Bilingualism and cognition. *Bilingualism: Language and Cognition*, **18**(1), 3–24.

Valian, V. (2015). Bilingualism and cognition: A focus on mechanisms. *Bilingualism: Language and Cognition*, **18**(1), 47–50.

Williams, K. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Wyche, K. F., Saegert, S. C., Adler, N. E., Bullock, H. E., Cauce, A. M., & Liu, W. M. (2019). *Report of the APA task force on socioeconomic status*. Washington, DC: American Psychological Association.

Zhao, T. C., & Kuhl, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proceedings of the National Academy of Sciences*, **113**(19), 5212–5217.

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