A Holistic Measure of Contextual and Individual Linguistic Diversity

The sociolinguistic context of language use contributes to individuals' linguistic knowledge. However, how it does so has yet to be fully conceptualised or quantitatively investigated within the language sciences. To meet this goal, a psychometrically sound evidence-based measure that captures important aspects of contextually influenced linguistic experience is necessary. In this paper we describe the development and validation of the Contextual and Individual Linguistic Diversity Questionnaire (CILD-Q). Exploratory factor analysis was conducted with data from 353 participants (62.9% South African, 37.1% UK, *M*age = 29.3, *SD* = 10.09). A three-factor solution best describes the structure of the CILD-Q: Multilingualism in Context (contextual use and societal practice of multiple languages within a community), Multilingualism in Practice (direct and indirect linguistic exchanges and conversational interaction), Linguistic Diversity Promotion (societal and governmental endorsement of linguistic variation). Item scores corresponding to these three factors showed sufficient reliability (α's > 0.80). The CILD-Q provides a novel and holistic manner to measure sociolinguistic diversity and can be used when measuring individuals' language experience within and across populations from differing sociolinguistic contexts.

Keywords: sociolinguistic context; contextual linguistic diversity; linguistic experience; language profile; CILD-Q; CLiP-Q

Introduction

There is little doubt that worldwide multilingualism is on the rise. Even so, it is obvious that some contexts are linguistically more diverse than others. This variation is a result of a multitude of factors, including contextual history, societal uptake, cultural and linguistic preservation, and political ideals. Sociolinguistic context may be a fundamental determinant of language knowledge and use, yet there remains a gap in quantitatively exploring or conceptualising *whether* and/or *how* one's sociolinguistic context facilitates linguistic gains (Bak, 2016). Generally, individual reports of language history, use, and proficiency are considered sufficient for language profiling (Anderson et al., 2018; Li et al., 2019; Marian et al., 2007), but this neglects the possible influence of *linguistic experience* within a social context on one's linguistic knowledge, including an awareness of this linguistic knowledge and ability. Given the plausibility of its effects on one's language development and awareness, the social context within which one is exposed to language should be a principled avenue of investigation.

Broadly, one's linguistic experience is fundamentally linked to their *sociolinguistic context* – the language environment within which speakers have spent most of their lives. Measuring one's linguistic experience provides a holistic overview of where the speaker is linguistically immersed, including how certain languages may be privileged over others, and the speakers' fluidity and inter-mingling of language use. This paper aims to develop and validate a psychometrically sound, evidence-based tool for measuring one's linguistic experience as it relates to their sociolinguistic context – the Contextual and Individual Linguistic Diversity Questionnaire (CILD-Q). Development of this questionnaire is important for a number of reasons. First, because it allows researchers to quantify and qualify language experience for monolingual, bilingual, and multilingual individuals thus adding sociolinguistic measures of linguistic

exposure that may be indirect and therefore may not qualify as 'input' in the usual sense of the word (Carroll, 2017). Second, because it can provide another, as yet under-researched, variable in the continuum of mono- to multi-lingualism which can contribute to an account of the heterogeneity found across so-called monolinguals and multilinguals in and beyond the Global North. Third, because a measure of the sociolinguistic context provided by this questionnaire could inform education policies about the challenges to be faced when models of language education deemed successful in one part of the world are assumed to be applicable to another. Such a measure is an integral element for any study aimed at illustrating that linguistic experience, and one's self-perception thereof, plays a fundamental role in the development of an individual's language repertoire.

We propose that linguistic experience is constituted by two features: (i) *active linguistic experience* – the direct use and communicative knowledge of a language, where one develops the ability to produce and consciously understand the language with a certain degree of proficiency; and (ii) *passive linguistic experience* – a novel concept which includes linguistic knowledge derived purely as a result of environmental cues mainly gained implicitly because of the context within which one or more languages are used. Whereas active linguistic experience is commonly used in language research, passive experience is possibly an integral element of one's linguistic profile, albeit not typically taken into account (but see Bice & Kroll, 2019; Kurkela et al., 2019; Tsimpli et al., 2020). In linguistically diverse contexts where numerous languages are spoken, displayed across the linguistic landscape (Gorter, 2013), and used interchangeably in daily exchanges, it is possible that at least some linguistic information from each of these languages may be cognitively integrated by people exposed to such information. Essentially, long-term experience within a sociolinguistic context may therefore

facilitate passive gains of linguistic knowledge over and above that which has been implicitly or explicitly gained through direct linguistic interactions during language development and use. Governmental and societal promotion of multilingualism can also facilitate active as well as passive linguistic gains in linguistic experience.

The way language is conceived of is continuously being updated with emerging evidence. Importantly, there is some evidence that supports the influence of passive experience. Some studies have found that, despite not retaining explicit knowledge of an exposed-to language into adulthood, childhood overhearers were better able than non-overhearers to learn phonological features of the overheard language later in life (Au et al., 2002, 2008; Knightly et al., 2003). Additionally, behavioural and neuroimaging research on functionally monolingual international adoptees point to the conclusion that even durationally minimal exposure to a language in childhood with no subsequent maintenance or conscious recollection of the language leads to perdurable linguistic and neural effects similarly expected in bilinguals who speak the 'lost' language (Oh et al., 2010; Pierce et al., 2014, 2015). Furthermore, language knowledge has been observed in the absence of deliberate learning and only via exposure alone (Akhtar et al., 2012; Gullberg et al., 2010), when participants are placed in an immersive learning setting rather than a non-immersive environment (Kroll et al., 2018; Morgan-Short et al., 2012; Pliatsikas et al., 2017), and via receptive input in heritage bilingual communities (Sherkina-Lieber, Pérez-Leroux, & Johns, 2011).

The aforementioned research suggests that even passive linguistic experience can have enduring effects on linguistic ability. Consequently, exposure to a linguistically diverse context cannot be downplayed when investigating overall linguistic experience even when interpersonal verbal communication may not require understanding, knowledge, and use of the totality of languages available in this context.

What is (sociolinguistic) context and how has it been investigated?

Crucial to our understanding of linguistic diversity is whether an individual is exposed to a multilingual or (predominantly) unilingual context both at an individual and societal level and how this affects her linguistic experience. For instance, relevant to the present research, a self-described monolingual English speaker from South Africa will be situated in a different sociolinguistic context to that of a self-described monolingual English speaker from the United Kingdom (UK).[1] In both countries English is the *lingua franca*, the medium of instruction for formal education, and is privileged in terms of its prominent use in business, media, and academia. Yet, since South Africa is host to 11 official languages (and numerous unofficial languages and dialects), the country's English-speakers are ubiquitously exposed to linguistic diversity in a way that differs from that of the English speakers in the UK. Although not every South African citizen may be exposed to all languages, linguistic heterogeneity within the country is largely unavoidable. For example, singing the South African national anthem alone employs the use of five of the official languages. Despite such cases, there is arguably no research that conceptualises or measures linguistic experience in relation to such sociolinguistic features. Of note, however, is a proposed metric of language entropy that captures the social diversity of language use derived from commonly collected language history data (Gullifer & Titone, 2020). Gaining insight into both individual and contextual linguistic descriptors can illuminate fundamental features of sociolinguistic diversity and provide an in-depth profile of language experience. Before we propose a way to do this, it is important that we understand how 'context' has been conceptualised in the literature.

Van Dijk (2008) suggests that in efforts to *describe* and especially *explain* some phenomenon in relation to its environment, we appeal to the notion of 'context'. A main application of contextualisation or contextualism[2] under this view has generally been

subsumed within a qualitative paradigm, using ethnography, discourse, and conversation analytic techniques as methods of analyses and natural talk as main data sources (ten Have, 2007; van Dijk, 1985, 2008). Another interactional method proposed to investigate context involves research into code-switching and borrowing which have garnered great interest as a means of conceptualising the bilingual experience (e.g., Grosjean, 1982; MacSwan 2014; Muysken, 2000). Linked to this is the Adaptive Control Hypothesis (ACH), which states that there are three types of interactional contexts that explain cognitive processing demands for bilingual speakers, namely the *single language*, *dual language*, and *dense code-switching* contexts (Green & Abutalebi, 2013; also see Abutalebi & Green, 2016). Single language refers to the use of one language in a particular setting and the use of another in a separate setting. Dual language is where different languages are typically produced by different speakers within the same context. Here, code-switching can occur in the conversation but not within a single utterance. Thirdly, dense code-switching is the fluid insertion of a speaker's languages in the course of an utterance and the adaptation of words from one language in the context of the other (also known as intrasentential code-switching). According to the ACH, variation in the context of an interactional exchange constrains the linguistic output and degree of language switching a bilingual can produce, thus uniquely impacting cognitive control. Language processing demands are then, in principle, greater for bilinguals, and accordingly, language selection shaping a bilingual's speech production confers a cognitive challenge that may be absent in monolingual brains (also see Blanco-Elorrieta & Pylkkänen, 2018). Cognitive adaptation is better understood, according to this hypothesis, when the context of language use is considered, as it is the interactional context that drives an adaptive response over and above a generalised bilingual advantage. While this hypothesis adds

to an understanding of context, it fails to acknowledge the sociolinguistic and cultural context within which the different languages are acquired and spoken, it is applicable only to bilingual contexts and speakers, and relies on instantaneous interactions of speech exchange, without considering any macro-level accounts of context.

The issues with the ACH are indicative of another trend within the literature: code-switching has predominantly been investigated as a bilingual phenomenon (e.g., Beatty-Martínez et al., 2020; Broersma et al., 2020; Carter et al., 2011) using natural conversations or sentences from corpora (Broersma & De Bot, 2006; Deuchar et al., 2018; Myers-Scotton, 2006) or experimental tasks usually requiring single word responses (Blanco-Elorrieta et al., 2018; Bobb & Wodniecka, 2013; Hernandez, 2009; Kleinman & Gollan, 2016) as analytic methods. Importantly, it is this characterisation of code-switching as a consequence of bilingual communication that has limited its scope and investigation. It need not only be a gathering of bilingual speakers that facilitate the interchange of language use, but rather, code-switching may arise so long as the sociolinguistic context is one of facilitation, regardless of the *lingualism status*[3] descriptors imposed. So far, little attention has been paid to code-switching within strictly speaking monolingual speakers exposed to linguistically diverse contexts (see MacSwan, 2017). This is particularly the case in South Africa, where linguistic amalgamation is dominant, even with self-reported monolingual speakers (Mesthrie, 2009). For instance, if a South African considers herself a monolingual English speaker, in the sense that English feels most comfortable for her to use across different modalities (speaking, listening, reading, writing), her passive linguistic experience may nevertheless include the use of other languages when conversing with linguistically similar or diverse people of the same sociolinguistic context. This type of amalgamated communication has been reported across numerous formal and informal settings,

including in advertising (Cowley, 2016), public hearings (Mabule, 2019), the classroom (Rose & Van Dulm, 2011), and townships (Finlayson & Slabbert, 2009). Thus, code-switching is not just a unique feature of bilingualism. Monolinguals exposed to linguistically diverse contexts may also switch seamlessly if afforded the opportunity to do so. The emphasis on bilingualism as the prominent code-switching criterion limits the bounds of language flexibility evident in interlocutors across the lingualism continuum and within specific contextual confounds. Critically, we need to reconsider the bounds of how we define communication as a feature of sociolinguistic context.

All the above indicates a tendency to merely describe one's immediate context, without noting a possible link between this context and other aspects of one's language profile. This results in the common approach of focusing on individual, internal linguistic differences and (if available) census data, rather than including the broader contextual milieu within which the individuals are situated. Kaushanskaya et al. (2019) note that 'self-reported proficiency [use and history] data alone CAN sometimes meet the needs of a particular study. However, a COMPREHENSIVE approach to assessing bilinguals is more likely to yield a reliable picture of bilinguals' language profiles' (p. 3). Importantly, it is the participants of study in relation to the environments within which they are linguistically exposed, culturally immersed, and have first-hand experience, that needs further exploration.

The present study

There is a need for a valid and reliable evidence-based measure that captures important aspects of contextual linguistic experience that can be widely implemented in and beyond the Global North. Accordingly, we developed the Contextual and Individual Linguistic Diversity Questionnaire (CILD-Q), as part of the larger Contextual Linguistic

Profile Questionnaire (CLiP-Q; see Methods), and psychometrically evaluated it using exploratory factor analysis. The investigation of the CILD-Q using this factor analysis is fundamental to further development and operationalising of the questionnaire. To our knowledge this is the first tool to directly measure contextual and individual linguistic diversity in addition to descriptors of language background, usage, and proficiency. Whereas such self-reports do provide valuable information about the linguistic repertoire of participants, they do not capture the full picture of how the sociolinguistic context adds to an individual's linguistic experience. To address concerns posed about the lack of universally applicable measures that quantify linguistic experience (Calvo et al., 2016), we demonstrate the construct validity and reliability of a new data collection tool that considers the effect of an increasingly diverse linguistic context on an individual's linguistic repertoire. We hypothesise that items measuring a single construct would group together, yielding the underlying factors that comprise contextual and individual linguistic diversity. We also assess the internal consistency reliability across the factor analysed scales of the CILD-Q.

Method

Participants

A sample of 353 participants completed the Contextual Linguistic Profile Questionnaire (CLiP-Q), of which 222 (62.9%) reported to have lived in South Africa and 131 (37.1%) in the UK for the longest period of their lives. Participants ranged in age from 18 to 59 years (M = 29.3, SD = 10.09). For those participants who reported some proficiency of a language other than English, 28 languages were represented. See Supplementary Material for additional demographic information.

Contextual Linguistic Profile Questionnaire (CLiP-Q)

The CLiP-Q is a holistic measure of contextual and individual linguistic diversity, comprising four sections, which together provide a comprehensive linguistic profile of participants across different sociolinguistic settings (see Figure 1 and refer to the Supplementary Material for an overview of the CLiP-Q). It was developed to compare adults from different sociolinguistic contexts; in this case South Africa and the UK. It is administered online in English and takes approximately 20 minutes to complete depending on the number of proficient languages reported. In what follows, we outline each of the CLiP-Q sections.

<Figure 1>

Demographic information

This section gathers information about where the participants have lived for the longest part of their lives, and other demographic information such as nationality, country and province/region of birth and current residence, age, gender, level of education, and employment status.

The Contextual and Individual Linguistic Diversity Questionnaire (CILD-Q)

The CILD-Q was developed to conceptualise and differentiate linguistic diversity both within the individual and as a feature of contextual exposure. It is designed in relation to where participants have lived for the longest period of time, with English as the reference language. The validated CILD-Q comprises 18 items (see Table 1) presented in a randomised order and scored on a five-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), with high scores reflecting higher exposure to different

languages from that of English (items 1-7), greater individual and conversational exposure to linguistic diversity (both as a factor of code-switching and borrowing; items 8-14), and higher governmental and societal encouragement of linguistic variation (items 15-18).

<Table 1>

Language history, use, and proficiency

This section was informed by existing questionnaires with items adapted from the Language Experience and Proficiency Questionnaire (Marian et al., 2007), the Language History Questionnaire (Li et al., 2019), and the Language and Social Background Questionnaire (Anderson et al., 2018), and was divided into two components. The first gathers information about general language background (e.g., home and written languages), exposure to English, and lingualism status self-classification. The second consists of questions about the participants' reported proficient languages (repeated for up to three languages). Participants indicate their most proficient language and the succeeding questions pertain to that language. For example, if a participant indicates that her most proficient language is Zulu, 'Zulu' is visible in the phrasing of succeeding questions, such as '*At what age did you begin acquiring Zulu?*' This section includes information for each language about average exposure; age milestones (e.g., acquisition, writing, use with friends); years of language use; level of ability in speaking, understanding, reading, and writing; extent of language use with interlocutors (e.g., family, strangers) and across activities (e.g., watching TV, listening to the radio/music); time engaged in activities (e.g., reading for school/work, writing messages to friends); duration of language used in different situations (e.g.,

thinking, expressing emotion); and degree of language association with culture (see Supplementary Material for descriptive statistics). This section concludes with an open-ended question where participants provide additional information about their language background and use. Reference to this information clarifies unclear findings. Since this section is a composite of questions commonly used in language research, a certain degree of construct and content validity can already be assumed.

Socioeconomic status

Participants indicate (yes/no) whether they have a list of assets or services in their household (e.g., motor vehicle, internet access). Other questions include number of contributors to annual household income, approximate annual household income, means of funding education, maternal and paternal level of education and occupation, and financial status (e.g., living from month-to-month, wealthy).[4]

Procedure and ethical considerations

Ethical clearance was obtained from the University of Cambridge Research Ethics Committee before this study commenced. The CLiP-Q was developed in successive stages, and items comprising the CILD-Q were developed by the authors as guided by theoretical and empirical research within the field of multilingualism. An initial pool of 30 items were created which were hypothesised to load across three constructs: general language exposure (including across the linguistic landscape and via communicative exchange), codeswitching and borrowing, and linguistic diversity promotion (see Results for the final factor structure and Supplementary Material for the original CILD-Q items). Care was taken to include a balanced number of items per factor (10 in each, equally split into positively and negatively worded items to avoid participant and acquiescence bias) and to ensure that items were clear, concise, and fit-for-purpose

before a factor analysis was conducted (Irwing & Hughes, 2018). To evaluate the content and face validity of the CLiP-Q, individuals across our network (including subject matter experts, test design experts, and representatives of participant samples) reviewed each of its sections, following which we addressed any comments and concerns raised, and made revisions where necessary until such time that a comprehensive language profile questionnaire was established. Once a final design was jointly approved, the CLiP-Q was transferred onto Qualtrics and distributed across the researchers' networks. Individuals wanting to participate did so voluntarily and without coercion. Informed consent was requested online via accessing the CLiP-Q along with an information sheet. Adherence to ethical guidelines and the Data Protection Act were observed throughout.

Data analysis

The focus of this paper was to assess the psychometric properties of the CILD-Q for use across differing sociolinguistic contexts. An exploratory factor analysis (EFA) was conducted in *R* (version 3.6.0; R Core Team, 2019) using the 'EFAutilities' (version 2.0.0; Zhang et al., 2019) and 'psych' (version 1.8.12; Revelle, 2019) packages with maximum likelihood estimation. An oblique geomin rotation, in which factors are not constrained to be orthogonal, was applied to the unrotated factor solution (Henson & Roberts, 2006; Osborne et al., 2008). An attractive feature of the EFAutilities package is that 95% confidence intervals are provided for rotated factor loadings, which allow for a more nuanced interpretation of point estimates.

Two stages are followed when computing an interpretable factor analysis with the aim of reducing variable complexity and assessing construct validity. Firstly, deciding on the optimal number of factors to extract, whereby an initial estimation of

item loadings based on correlation or covariance matrices are considered as indicators of the number of factors required to account for common variance. Secondly, a factor rotation method is applied, which involves a mathematical transformation of unrotated factor loadings for ease of factor interpretability (Gorsuch, 1983; Izquierdo et al., 2014). It has been urged to use 'multiple criteria and reasoned reflection' (Henson & Roberts, 2006, p. 399) when selecting the most appropriate number of factors to retain, and that 'the simultaneous use of multiple decision rules is appropriate and often desirable' (Thompson & Daniel, 1996, p. 200; also see Auerswald & Moshagen, 2019). We used a scree plot, parallel analysis, Velicer's minimum average partial (MAP) test, and the proposed a priori theory (Cattell, 1966; Horn, 1965; Lorenzo-Seva et al., 2011; Velicer, 1976). Once a factor solution was established, internal consistency reliability using Cronbach alpha was used to evaluate the reliability of each scale score and the overall CILD-Q (Revelle & Condon, 2018).

Results

Preliminary results and exploratory factor analysis

The EFA was conducted on a sample of 353 participants. Communality coefficients (presented in Table 3) are mostly greater than 0.50 and therefore using factor analysis with the current sample size was justified (MacCallum et al., 1999). Data are reported for the factor analysed solution, which was reached after two stages of item removal.[5]

Two measures of sampling adequacy were reviewed prior to conducting the factor analysis. The Kaiser–Meyer–Olkin (KMO) criterion indicates if there is sufficient common variance between items to warrant factor analysis (Kaiser, 1974) and the Bartlett's test of sphericity assesses the overall significance of the correlations within the matrix (Bartlett, 1937). The KMO value was 0.92, which is considered as a

marvellous adequacy of correlations and Bartlett's test of sphericity was significant, $\chi^2(153) = 4434.28$, $p < 0.001$, indicating that factor analysis of the correlation matrix was suitable. Univariate descriptive statistics for the 18 CILD-Q items are presented in Table 2.

<Table 2>

Construct validity of the CILD-Q

Selection of factors were based on an interpretation of the scree plot, parallel analysis, Velicer's MAP test, and the proposed a priori theory. The scree plot and parallel analysis suggested four factors (see Figure 2), while MAP suggested three factors, and following a thorough inspection of both a three- and four-factor solution, we identified that the four-factor solution led to a separation of items that clustered together when restricted to a three-factor solution as well as increased the number of cross-loadings. Further inspection of the items content, with the consideration of multilingualism theory, led us to appraise the three-factor solution, with its integrated and theoretically sound factor, as more appropriate.

Variable complexity was also considered in the interpretation of factors. Sass and Schmitt (2010) define variable complexity as the number of nonzero elements in a factor pattern matrix row and it is evaluated in relation to cross-loading magnitudes. The pattern matrix after rotation can be seen in Table 3. From the factor pattern matrix, it is evident that no cross-loadings exceeded 0.30 across all items, with the highest cross-loading reaching 0.30, while the majority of other loadings reaching a magnitude of below 0.10. The complexity structure of the data is therefore considered *approximate simple structure*.

<Figure 2>

Based on the above extraction and complexity evidence as well as the theoretical construction guiding the development of this questionnaire, a three-factor solution was evaluated as best describing the structure of the CILD-Q.

<Table 3>

Inspection of the clustering items that make up the three factors suggests that Factor 1 represents *Multilingualism in Context*, Factor 2 represents *Multilingualism in Practice*, and Factor 3 represents *Linguistic Diversity Promotion*.

The eigenvalues and (percentage of common variance) for the three factors were: (1) Multilingualism in Context (seven items) = 4.78 (43.1%), (2) Multilingualism in Practice (seven items) = 3.55 (32.0%), and (3) Linguistic Diversity Promotion (four items) = 2.76 (24.9%). The total common variance explained by the structure is 61.6% with all three factors correlating positively with one another. From these results, the construct validity of the 18-item CILD-Q is demonstrated.

Descriptive statistics and reliability of the CILD-Q

Means, standard deviations, and internal consistency reliability of each scale and overall CILD-Q score are presented in Table 4. Alpha coefficients range from 0.85 to 0.93, with an overall CILD-Q reliability coefficient of 0.91, suggesting that the CILD-Q is a reliable measure of contextual and individual linguistic diversity.

<Table 4>

Discussion

In this study we aimed to develop a measure of linguistic experience with a specific focus on contextual and individual linguistic diversity in the English-speaking contexts of South Africa and the UK. Though we intend the CLiP-Q to be employed as needed beyond these contexts. Due to linguistic diversity in South Africa and increasingly so in the UK, multilingualism is exceedingly prevalent and important. It is the norm rather than the exception that people are exposed to and can oftentimes speak more than one language. This fosters social, educational, and cognitive implications and is a key aspect to understanding linguistic development across these contexts. To our knowledge, this is the first attempt to create a measure that evaluates linguistic experience by evaluating diversity both as a feature of the individual's language profile as well as the sociolinguistic context within which the individual is situated. Our questionnaire can provide a quantitative indication of active and passive language experience for individuals across the lingualism continuum and located in various parts of the world, in a way that has not been conceived of before. We evaluated the psychometric soundness of the CILD-Q and from the exploratory analysis, yielded a three-factor solution that was found to be valid and reliable. The factors, in order of variance explained, are operationalised as follows:

Multilingualism in Context. This scale is thought to measure the contextual use and societal practice of multiple languages in addition to the dominant language within a context (in this case English). Specifically, this scale refers to linguistic experience as a feature of linguistic diversity in the social context. It is the culmination of linguistic diversity that occurs at a societal level, where multilingualism is facilitated (or impeded) via exposure to oral and written content in one's immediate environment. This could be

through the media, social network platforms, signage information and input, and a general uptake and usage of language variety within a sociolinguistic community.

Multilingualism in Practice. Our second scale is operationalised as a measure of direct and indirect linguistic exchanges and conversational exposure. That is, individual exposure to linguistic diversity as a feature of spoken engagement that one is either directly or indirectly involved in. This can occur via direct communication between speakers or as a circumstance of overheard or ancillary exchanges, such as listening to a conversation between strangers as a passer-by. Aspects of code-switching and borrowing can encompass both this and the Multilingualism in Context scale in unique ways. For the former, code-switching and borrowing emerge as a feature of personal communication between members within a society, whereas for the latter, code-switching and borrowing manifest generally as a common language practice within a society.

Linguistic Diversity Promotion. Finally, our third scale is theorised to measure societal and governmental endorsement and promotion of linguistic variety within the context. Uptake of a particular language(s) is largely dependent on whether there is a positive motivation to do so within the socio-political climate of a context.

Taken together, the inclusion of the three aspects of linguistic diversity (in the context, in the individual's linguistic experience, and in the societal attitudes) provides a novel and holistic way to measure sociolinguistic diversity. In this way, it promotes our understanding of linguistic experience, by incorporating language practices in the individual and in the context within which the individual is linguistically situated. The CLiP-Q can capture variation among adults with different degrees of linguistic experience. The CILD-Q can be used in isolation and should be adapted specifically to meet the needs of the population and sociolinguistic context under investigation.

Although the questionnaire was designed with English as the reference language, we encourage adaptation and translation of the CILD-Q into any language or dialect appropriate to meet the research needs of a project.

Implications for further research on bilingualism

The linguistic representation of a social context influences the people living within such a context. Although 'context' has largely been ignored or taken for granted within much of the language literature, an acknowledgement of its importance and steady transition toward a more integrated method has been suggested (Bak, 2016; Green & Abutalebi, 2013; Luk & Bialystok, 2013). Building toward this goal, a particularly important area where this research can advance academic insight is bilingualism/multilingualism research, where debates surrounding the 'bilingual advantage' have yet to be addressed in an innovative methodological manner. It is proposed that because bilinguals must manage and monitor the competition of multiple languages simultaneously, this process facilitates a cognitive advantage that is translated to tasks measuring similar cognitive abilities (Bialystok, 2001), as well as other domains not limited only to cognition (e.g., novel word learning: Kaushanskaya & Marian, 2009; phonetic learning: Antoniou et al., 2015; working memory: Cockcroft et al., 2019). On the one hand, there are studies that support the bilingual advantage specifically in domain general executive function tasks, while other studies dispute that bilingualism is in itself a condition for which advanced executive control emerges (for reviews see Adesope et al., 2010; Bialystok et al., 2009; Bialystok, 2017; Bright & Filippi, 2019; Paap et al., 2015; van den Noort et al., 2019). While the point of this paper is not to delve into the larger debate at hand, we propose that the sociolinguistic context to which a population is exposed is an important aspect as to why some of these differences may emerge. Conceptualising and investigating

language use as a function of context, in combination with direct assessment of a speaker's abilities in language and cognition, could lead to a deeper understanding of the contradictory evidence within bilingualism research. Understanding speakers' diverse linguistic profiles as a factor of their contextual linguistic experience will foster a nuanced and holistic account of the aspects that may contribute to these differences.

Takahesu Tabori et al. (2018) claim that 'simply knowing whether individuals are immersed in the first or second language and whether they are proficient or not, is not sufficient' (p. 4). Likewise, Anderson et al. (2018) claim that:

> most research on bilingualism uses some form of self-report questionnaire to gather information relevant to this designation, but the design and interpretation of such instruments are vague because bilingualism is a multifaceted experience shaped by social, individual, and contextual factors (p. 250).

The language science community have conclusively advocated for novel approaches that can advance an understanding of linguistic knowledge across differing populations – the CILD-Q (as part of the larger CLiP-Q project) is one such attempt at heeding their call.

Limitations and recommendations for future research

Since this research is the first of its kind, one of the main caveats is the lack of exclusion criteria confinement. The CLiP-Q was accessible to all South Africans and people from the UK who were at least 18 years old. Such a large population pool may affect ecological validity. However, equally imperative was the aim to attain as large a representative sample as possible from each group in order to ensure a fair representation of speakers within the different sociolinguistic contexts. It is recommended that future research narrow the target population criteria (including age range, regional specificity, etc.) such that a clear target population be established

beforehand. Building on from this, it is necessary that the CLiP-Q be sampled on different linguistically diverse groups from various language contexts so as to ensure its validity across other groups and contexts. With the results presented, external evaluation of the CLiP-Q and specifically the CILD-Q in subsequent studies is encouraged. In addition, further statistical methods should be applied to address other psychometric properties of the measure, such as using confirmatory factor analysis and an Item Response Theory framework to ensure appropriate item analysis. Another limitation concerns participants needing access to an electronic device and an internet connection in order to complete the CLiP-Q. This limitation has important implications about who the sample represents and their potential socioeconomic advantage. These results are tentative and further studies using the CILD-Q are required to replicate these initial findings.

Conclusion

This exploratory analysis of the emerging factors of the CILD-Q provide important insights into the underlying structure of contextual linguistic diversity. A questionnaire of this nature may have a pertinent role to play in the comparison of linguistically diverse groups from differing linguistic contexts. We propose that the CILD-Q should be used in research settings and adapted accordingly whenever there is a comparison of linguistically diverse groups from different contexts. The context of reference can be adjusted as needed based on the study's priorities and aims, and the reference language need not be English but any language(s) under investigation.

This paper illustrates the importance of contextual linguistic diversity as an imperative measure of holistic linguistic diversity. One of the imperatives for a meaningful understanding of human interaction is to consider the unified relationship

between language, context, and society. Prioritising contextual linguistic experience as an outside influence on linguistic knowledge, in supplementation to individual assessments of speakers through direct measures, is perhaps the key to a progressive grasp of language differences that are thought to give rise to cognitive adaptations. By considering a novel approach of language experience across differing sociolinguistic contexts, we have been able to develop a comprehensive account of contextual linguistic diversity and illustrate the relevance of incorporating it into language research.

Acknowledgements

We would like to thank Brandon Morgan for his statistical consultation and review of the manuscript, Adrian Erasmus and Ariane Hanemaayer for their insightful comments and review of an earlier draft of the manuscript, and Ianthi Tsimpli's research team (James Algie, Anusha Balasubramanian, Alex Cairncross, Christiana Christodoulou, Anthony Li, Ruijun Ma, Samantha Sie, Andromachi Tsoukala, Margreet Vogelzang, Jodie Webber, Jinxing Yue) for their participation in the review of the CLiP-Q items. A further thanks is extended to all the participants who took the time to take part in this study.

References

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition, 19*(4), 689–698. doi:10.1017/S1366728916000225

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. doi:10.3102/0034654310368803

Akhtar, N., Menjivar, J., Hoicka, E., & Sabbagh, M. A. (2012). Learning foreign labels from a foreign speaker: The role of (limited) exposure to a second language. *Journal of Child Language, 39*(5), 1135–1149. doi:10.1017/S0305000911000481

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. doi:10.3758/s13428-017-0867-9

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition, 18*(4), 683–695. doi:10.1017/S1366728914000777

Au, T. K.-F., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science, 13*(3), 238–243.

Au, T. K.-F., Oh, J. S., Knightly, L. M., Jun, S.-A., & Romo, L. F. (2008). Salvaging a childhood language. *Journal of Memory and Language, 58*(4), 998–1011. https://doi.org/10.1016/j.jml.2007.11.001.

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 469–491. http://dx.doi.org/10.1037/met0000200

Bak, T. H. (2016). Cooking pasta in La Paz: Bilingualism, bias and the replication crisis. *Linguistic Approaches to Bilingualism, 6*(5), 699–717. doi:10.1075/lab.16002.bak

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society, Series A, 160*, 268–282.

Beatty-Martínez, A. L., Navarro-Torres, C. A., & Dussias, P. E. (2020). Codeswitching: A bilingual toolkit for opportunistic speech planning. *Frontiers in Psychology, 11*, 1699. doi:10.3389/fpsyg.2020.01699

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143*(3), 233–262. http://dx.doi.org/10.1037/bul0000099

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, 10*(3), 89–129. doi:10.1177/1529100610387084

Bice, K., & Kroll, J. F. (2019). English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language, 196*, 104644. https://doi.org/10.1016/j.bandl.2019.104644

Blanco-Elorrieta, E., Emmorey, K., & Pylkkänen, L. (2018). Language switching decomposed through MEG and evidence from bimodal bilinguals. *Proceedings of the National Academy of Sciences, 115*(39), 9708–9713. doi/10.1073/pnas.1809779115

Blanco-Elorrieta, E., & Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences, 22*(12), 1117–1126. https://doi.org/10.1016/j.tics.2018.10.001

Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology, 25*(5), 568–585. http://dx.doi.org/10.1080/20445911.2013.792822

Bright, P., & Filippi, R. (2019). Perspectives on the "bilingual advantage": Challenges and opportunities. *Frontiers in Psychology, 10*, 1–3. doi:10.3389/fpsyg.2019.01346

Broersma, M., Carter, D., Donnelly, K., & Konopka, A. (2020). Triggered codeswitching: Lexical processing and conversational dynamics. *Bilingualism: Language and Cognition, 23*(2), 295–308. https://doi.org/10.1017/S1366728919000014

Broersma, M., & De Bot, K. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition, 9*(1), 1–13.

Calvo, N., García, A. M., Manoiloff, L., & Ibáñez, A. (2016). Bilingualism and cognitive reserve: A critical overview and a plea for methodological innovations. *Frontiers in Aging Neuroscience, 7*(249), 1–17. doi:10.3389/fnagi.2015.00249

Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism: Language and Cognition, 20*(1), 3–16. doi:10.1017/S1366728915000863

Carter, D., Deuchar, M., Davies, P., & del Carmen Parafita Couto, M. (2011). A systematic comparison of factors affecting the choice of matrix language in three

bilingual communities. *Journal of Language Contact, 4*(2), 153–183. doi:10.1163/187740911X592808

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276.

Cockcroft, K., Wigdorowitz, M., & Liversage, L. (2019). A multilingual advantage in the components of working memory. *Bilingualism: Language and Cognition, 22*(1), 15–29. doi:10.1017/S1366728917000475

Cowley, J. (2016). *Code-switching as a persuasive device in South African advertising* [Unpublished master's thesis]. Stellenbosch University.

Department for Education. (2014, December). *The National Curriculum in England – Framework Document*. Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/381344/Master_final_national_curriculum_28_Nov.pdf

Deuchar, M., Webb-Davies, P., & Donnelly, K. (2018). *Building and using the Siarad Corpus: Bilingual conversations in Welsh and English.* John Benjamins Publishing Company.

Finlayson, R., & Slabbert, S. (2009). "We just mix": code switching in a South African township. *International Journal of the Sociology of Language, 125*(1), 65–98. https://doi.org/10.1515/ijsl.1997.125.65

Gorsuch, R. L. (1983). *Factor analysis*. Lawrence Erlbaum Associates.

Gorter, D. (2013). Linguistic landscapes in a multilingual world. *Annual Review of Applied Linguistics, 33*, 190–212. https://doi.org/10.1017/S0267190513000020

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. http://dx.doi.org/10.1080/20445911.2013.796377

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Harvard University Press.

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning, 60*(2), 5–24. https://doi.org/10.1111/j.1467-9922.2010.00598.x

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition, 23*, 283–294. https://doi.org/10.1017/S1366728919000026

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393–416. doi:10.1177/0013164405282485

Hernandez, A. E. (2009). Language switching in the bilingual brain: What's next? *Brain and Language, 109*(2–3), 133–140. doi:10.1016/j.bandl.2008.12.005

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185.

Irwing, P., & Hughes, D. J. (2018). Test development. In P. Irwing, T. Booth, & D. J Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 1–47). John Wiley & Sons Ltd.

Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema, 26*(3), 395–400. doi:10.7334/psicothema2013.349

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36.

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. https://doi.org/10.1017/S1366728919000038

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin and Review, 16*(4), 705–710. doi:10.3758/PBR.16.4.705

Kleinman, D., & Gollan, T. H. (2016). Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological Science, 27*(5), 700–714. doi:10.1177/0956797616634633

Knightly, L. M., Jun, S.-A., Oh, J. S., & Au, T. K. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America, 114*(1), 465–474. doi:10.1121/1.1577560

Kroll, J. F., Dussias, P. E., & Bajo, M. T. (2018). Language use across international contexts: shaping the minds of L2 speakers. *Annual Review of Applied Linguistics, 38*, 60–79. doi: 10.1017/S0267190518000119

Kurkela, J. L. O., Hämäläinen, J. A., Leppänen, P. H. T., Shu, H., & Astikainen, P. (2019). Passive exposure to speech sounds modifies change detection brain responses in adults. *NeuroImage, 188*, 208–216. https://doi.org/10.1016/j.neuroimage.2018.12.010

Li, P., Zhang, F., Yu, A., & Zhao, X. (2019). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 1–7. https://doi.org/10.1017/S1366728918001153

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research, 46*(2), 340–364. doi:10.1080/00273171.2011.564527

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. doi:10.1080/20445911.2013.795574

Mabule, D. R. (2019). Code-switching as a form of communication: Insights from a South African public hearing discussion. *South African Journal of African Languages, 39*(1), 67–75. doi:10.1080/02572117.2019.1572324

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99.

MacSwan, J. (2014). *Grammatical theory and bilingual codeswitching*. MIT Press.

MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal*, *54*(1), 167–201. doi:10.3102/0002831216683935

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. doi:10.1044/1092-4388(2007/067)

Mesthrie, R. (2009). Sociolinguistics in South Africa: A critical overview of current research. In M. J. Ball (Ed.), *The Routledge handbook of sociolinguistics around the world* (pp. 187–202). Routledge.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*(4), 933–947. https://doi.org/10.1162/jocn_a_00119

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press.

Myers-Scotton, C. (2006). Natural codeswitching knocks on the laboratory door. *Bilingualism: Language and Cognition, 9*(2), 203–312. doi:10.1017/S1366728906002549

Oh, J., Au, T., & Jun, S. (2010). Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language, 37*(5), 1123–1132. doi:10.1017/S0305000909990286

Osborne, J. W., Costello, A. B., & Kellow, J. T. (2008). Best practices in exploratory factor analysis. In O. Osborne (Ed.), *Best practices in quantitative methods* (pp. 86–99). SAGE Publications.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265–278. http://dx.doi.org/10.1016/j.cortex.2015.04.014

Pierce, L. J., Chen, J.-K., Delcenserie, A., Genesee, F., & Klein, D. (2015). Past experience shapes ongoing neural patterns for language. *Nature Communications, 6*, 10073. https://doi.org/10.1038/ncomms10073.

Pierce, L. J., Klein, D., Chen, J.-K., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences of the United States of America, 111*(48), https://doi.org/10.1073/pnas.1409411111.

Pliatsikas, C., DeLuca, V., Moschopoulou, E., & Saddy, J. D. (2017). Immersive bilingualism reshapes the core of the brain. *Brain Structure and Function, 222*, 1785–1795. DOI 10.1007/s00429-016-1307-9

R Core Team. (2019). *R: The R project for statistical computing*. R Core Team. https://www.r-project.org/

Revelle, W. (2019). *Package "psych". R package version 1.8.12*. R Core Team. https://CRAN.R-project.org/package=psych

Revelle, W., & Condon, D. M. (2018). Reliability. In P. Irwing, T. Booth, & D. J Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 709–749). John Wiley & Sons Ltd.

Rose, S., & Van Dulm, O. (2011). Functions of code switching in multilingual classrooms. *Per Linguam, 22*(2), 1–13. http://dx.doi.org/10.5785/22-2-63

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45*(1), 73–103. doi:10.1080/00273170903504810

Sherkina-Lieber, M., Perez-Leroux, A., & Johns, A. (2011). Grammar without speech production: The case of Labrador Inuttitut heritage receptive bilinguals. *Bilingualism: Language and Cognition, 14*(3), 301-317. doi:10.1017/S1366728910000210

Takahesu Tabori, A. A., Mech, E. N., & Atagi, N. (2018). Exploiting language variation to better understand the cognitive consequences of bilingualism. *Frontiers in Psychology, 9*, 1–7. doi:10.3389/fpsyg.2018.01686

ten Have, P. (2007). *Doing conversation analysis* (2nd ed.). SAGE.

The British Academy. (2019). *Languages in the UK: A call for action*. The British Academy. https://www.thebritishacademy.ac.uk/sites/default/files/Languages-UK-2019-academies-statement.pdf

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*(2), 197–208.

Tsimpli, I. M., Vogelzang, M., Balasubramanian, A., Marinis, T., Alladi, S., Reddy, A., & Panda, M. (2020). Linguistic diversity, multilingualism, and cognitive skills: A study of disadvantaged children in India. *Languages, 5*, 10. doi:10.3390/languages5010010

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), 27. doi:10.3390/bs9030027

van Dijk, T. A. (1985). *Handbook of discourse analysis: Discourse analysis in society*. Academic Press.

van Dijk, T. A. (2008). *Discourse and context: A sociocognitive approach*. Cambridge University Press.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*(3), 321–327.

Zhang, G., Jiang, G., Hattori, M., & Trichtinger, L. (2019). *Package "EFAutilities". R package version 2.0.0*. R Core Team. https://CRAN.R-project.org/package=EFAutilities

---

[1] The UK appears to be less linguistically diverse in both number of speakers of multiple languages and policy (Department for Education, 2014; The British Academy, 2019).

[2] For a review and comparison, see van Dijk (2008).

[3] We use this term to describe monolinguals, bilinguals, and multilinguals across the continuum of language knowledge.

[4] Not all data from the CLiP-Q is presented in this paper due to space limitations and a focus on addressing the primary research aim.

[5] Items that did not have a meaningful or large enough loading (< 0.40), had cross-loadings, were more convoluted and lengthy, or did not have the largest loading in the pair of positively and negatively worded items were removed from the item pool after the first and second analyses. An initial pool comprised of 30 items and, following the EFA, was reduced to 18 items that best represent the internal structure of the latent construct.