

Sociolinguistic context matters: Exploring differences in contextual linguistic diversity in South Africa and England

Mandy Wigdorowitz^{a,b*}, Ana I. Pérez^c, and Ianthi M. Tsimpli^a

^aTheoretical and Applied Linguistics, University of Cambridge, Cambridge, United Kingdom; ^bDepartment of Psychology, University of Johannesburg, Johannesburg, South Africa; ^cDepartment of Experimental Psychology, University of Granada, Granada, Spain

*Corresponding author: Mandy Wigdorowitz, mw738@cam.ac.uk, St John's College, St John's Street, Cambridge CB2 1TP

Competing interests: The authors declare none

Abstract

Individual reports of language history, use, and proficiency are generally considered sufficient for language profiling. Yet, these variables alone neglect the contribution of contextual linguistic diversity to one's overall language repertoire. In this study we used the Contextual Linguistic Profile Questionnaire to evaluate whether there is a difference in contextual linguistic diversity between participants across the linguistically dissimilar contexts of South Africa and England. We further assessed whether self-reported lingualism status groups (monolinguals, bilinguals, multilinguals) scored differently on contextual linguistic diversity to evaluate the utility and uniformity of categorical labels across varying contexts, and investigated how codeswitching and socio-economic status contributed to these effects. Our results demonstrated that contextual linguistic diversity differs between nations: South Africans score higher, promotion of multilingualism is dependent on socio-economic status only in England, lingualism status is not contextually comparable, and codeswitching accounts for linguistic features of South Africans.

Keywords: CILD-Q; CLiP-Q; codeswitching; contextual linguistic diversity; language profiling; lingualism status; socio-economic status

Introduction

A growing body of literature has highlighted the impact of context on bilingual language use (Beatty-Martínez, Valdés Kroff, & Dussias, 2018; Kreiner & Degani, 2015; Montrul, 2015) and its effects on cognition more broadly (Abutalebi & Green, 2016; Freedman et al., 2014). It has been well documented that L2 learning imposes almost immediate and long-term changes to the L1, irrespective of whether a second language is learned early or late in life (Bice & Kroll, 2015; Chang, 2012; Schmid, 2013; Linck, Kroll, & Sunderman, 2009). However, fewer studies have addressed whether the sociolinguistic context from which speakers are drawn contributes to their language repertoire. This area of research is extended across the language continuum to speakers who may not consider themselves to be bilingual or multilingual in the usual sense of the word, but who may be immersed in a linguistically diverse context and gain linguistic information from their environment, without necessarily having an acute awareness of such gains.

Contextual diversity can include instances of interpersonal communication where it is not necessary for an interlocutor to attain full comprehension, knowledge, or use of the totality of languages available within their specific context, but where mutual understanding and recognition of meaning may nevertheless be attained. Such situations are largely evident in countries with a colonial past, where English is not always a principle mother tongue, yet occupies a place of prestige and socio-political power (Tsimpli et al., 2019). Accordingly, language knowledge can include linguistic information via either active or passive linguistic experience (Wigdorowitz, Pérez, & Tsimpli, 2020). *Active linguistic experience* is the direct use, production, and upkeep of language(s) that one employs in their regular communicative endeavours, where the speaker has a conscious representation and has developed some degree of proficiency in each of their languages. In contrast, *passive linguistic experience* is the summation of linguistic knowledge as a consequence of implicit and contextual linguistic

exposure and input that is derived from the sociolinguistic environment within which the speaker is situated. Passive linguistic experience is more commonly observed in multilingual societies and settings, where language information is mainly gained implicitly as a consequence of the sociolinguistic milieu.

While active linguistic experience is the common model used in language research, passive experience is steadily gaining recognition as an integral feature of one's language repertoire. For example, ambient exposure to multilingualism in a more multilingual society (southern California compared to central Pennsylvania) predisposes the brain of monolingual adults to learning another language (Bice & Kroll, 2019), while adults who are passively exposed to foreign speech sounds manifest an enhancement of auditory discrimination (Kurkela, Hämäläinen, Leppänen, Shu, & Astikainen, 2019). In fact, this learning effect seems to extend beyond the auditory domain, as it has been found that primary school children in India coming from monolingual households show an advantage in fluid intelligence as an effect of sociolinguistic diversity in their school and community settings (Tsimpli et al., 2020). Moreover, extensive literature has demonstrated that language knowledge is observed in the absence of deliberate learning and, possibly via limited language exposure (Akhtar, Menjivar, Hoicka, & Sabbagh, 2012; Gullberg, Roberts, Dimroth, Veroude, & Indefrey, 2010; Oh et al., 2020), when participants receive linguistic input from an immersive learning setting rather than a non-immersive one (Kroll et al., 2018; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Pliatsikas, DeLuca, Moschopoulou, & Saddy, 2017), and via receptive input in heritage bilingual communities (Sherkina-Lieber, Pérez-Leroux, & Johns, 2011).

Furthermore, the effects of passive exposure also appear to endure over one's lifespan. For instance, Au and colleagues found that childhood overhearers (adults who consistently overheard a language as children) were able to learn native-like phonological features of their

overheard language despite not retaining explicit knowledge or awareness of the exposed-to language (Au, Knightly, Jun, & Oh, 2002; Au, Oh, Knightly, Jun, & Romo, 2008; Knightly, Jun, Oh, & Au, 2003). Additionally, behavioural and neuroimaging research on functionally monolingual international adoptees allude that even durationally minimal exposure to a language in childhood with no subsequent maintenance or conscious recollection of the language leads to long-term linguistic and neural effects similarly expected in bilinguals who speak the ‘lost’ language (Oh, Au, & Jun, 2010; Pierce, Chen, Delcenserie, Genesee, & Klein, 2015). Overall, this evidence indicates that passive linguistic experience has an important effect on one’s language repertoire.

In light of these findings, it is possible that in linguistically diverse, multilingual contexts, where various languages are spoken and displayed across the linguistic landscape (Gorter, 2006), at least some linguistic information is cognitively integrated by the people situated within and exposed to such contexts. *‘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’* (Wigdorowitz et al., 2020, p. 2). Consequently, exposure to a linguistically diverse context over a substantial period of time is imperative to one’s language repertoire, particularly when exploring the holistic linguistic experience of speakers across the continuum of language knowledge.

Context and the lingualism status spectrum

Another area of importance in language research regards the classification of individuals into respective language groups, such as the commonly used descriptors of monolingualism, bilingualism, and multilingualism. We refer to these linguistic descriptors as an individual’s *lingualism status* since these describe some continuous stature of an individual’s language

experience. Yet, an obvious caveat observed from the accumulated literature is the lack of clarity and regularity in conceptualising the lingualism status of participants and the often-imposed categorical grouping of participants into one of these groups (Surrain & Luk, 2019). However, we propose that a categorical classification of one's lingualism status may pose unique problems to the conception of the individual's language knowledge and experience, given the limitations of language experience imposed by one linguistic label. It is therefore necessary to evaluate whether speakers are homogeneous in their contextual language exposure if they classify themselves with the same linguistic label because, under the conventional view, we would expect monolinguals (extended to bilinguals and multilinguals) from one context to be largely equivalent to monolinguals from another. Yet, are monolinguals in multilingual societies equal to (i) bilinguals/multilinguals within the same sociolinguistic context, or (ii) monolinguals in less linguistically diverse contexts? It is only when sociolinguistic diversity and passive linguistic exposure is added to an epistemology of language experience that the prospect of greater understanding about the confounds of one's language profile is real.

As noted, if monolingualism is a unanimous experience, then it is assumed that those who identify or are classified as monolingual have homogeneous and stable linguistic knowledge. However, such a restricted classification often fails to take contextual linguistic exposure and classificatory norms into account. The categorisation of participants into boxed language groups, whether self-reported or designated, may therefore conceal nuances of language knowledge and experience that could alter research findings across various domains. For instance, studies have found differences in electrophysiological responses to language processing across seemingly similar monolingual groups differing in contextual linguistic exposure (Bice & Kroll, 2019) and proficiency (Pakulak & Neville, 2010; Tanner & Van Hell, 2014). In addition, the situation is further complicated because different

linguistic contexts (whether across countries, communities, or at the individual level) employ unique standards and ideologies as to what counts as monolingualism, bilingualism, and multilingualism, as well as how language knowledge may be integrated as a result of the general linguistic milieu. As is the binary status quo, bilinguals are usually compared to monolingual counterpart “control” groups to assess whether bilingualism has an effect on X behavioural, linguistic, or cognitive outcome. Moreover, the majority of bilingualism research has been conducted within Global North, Western, or Anglosphere contexts, where conceptions of language knowledge and use may have particular meanings distinct from those from other more linguistically diverse contexts outside of the Global North.

In a study exploring the consequences of context on language use, Beatty-Martínez et al. (2019) compared three groups of ostensibly equivalent, highly proficient Spanish-English bilinguals, situated across three contexts that differed in language practice on lexical production (category verbal fluency and picture naming) and proactive/reactive inhibitory control. One group lived in Granada, Spain, where Spanish (majority language) and English are mainly used separately and across specific domains (e.g., Spanish at home, English at school/work, little-to-no codeswitching). The second group lived in San Juan, Puerto Rico, where Spanish and English are more integrated and there is greater flexibility of language choice across domains and more opportunities to codeswitch. The last group consisted of participants born and raised in Spanish-speaking environments who immigrated to the United States during childhood or adolescence and were studying at State College, Pennsylvania. Although English is the dominant language used across domains in this State, there are varied but limited opportunities for Spanish use and codeswitching. Overall, they found that lexical access and how it relates to cognitive control greatly depends on the language practices and demands of the linguistic context. Specifically, accuracy on the picture naming task was modulated by inhibitory control only for the immersed bilinguals in the United States but not

the other two groups, suggesting that this group of bilinguals needed to actively monitor their context prior to speaking, as their language choice was facilitated in relation to the language of their interlocutors, which had habitually become English. The authors acknowledged that most studies would normally aggregate such seemingly similar speakers into a single Spanish-English bilingual group, whereby this approach could lead to ‘*a failure to characterize the complexity associated with the context of language use*’ (Beatty-Martínez et al., 2019, p. 15).

Although much work has been done to describe and quantify language experience, there is little uniformity or standard method as to how and by what measure this should be captured. Furthermore, assessment efforts have largely amassed around children’s linguistic knowledge, with far fewer questionnaires available for adults (for a review see Kaščelan, 2020). Given the expansive interest in multilingualism research, the goal should be attaining language knowledge about the participants that meets a best practice standard, and we propose that this includes a measure that captures sociolinguistic experience along with history, usage, and proficiency indicators. In order to appropriately characterise speakers within dynamic language contexts, it is clear that we need to consider their diverse experiences with respect to contextual language use (Bice & Kroll, 2019; Kroll, Dussias, & Bajo, 2018; Tsimpli et al., 2020; Wigdorowitz et al., 2020). Having a holistic measure of language experience that captures information about sociolinguistic context can, accordingly, indicate where the speaker is linguistically immersed, including how certain languages may be privileged over others, and the flexibility, codeswitching, and exchange of language practices.

One recent questionnaire has been designed to evaluate both individual and contextual linguistic diversity as distinct features comprising the language profile: The Contextual and Individual Linguistic Diversity Questionnaire (CILD-Q; Wigdorowitz et al., 2020). The

CILD-Q assesses contextual linguistic diversity by three scales: a) *Multilingualism in Context*, which encompasses the contextual use, societal practice, and community language norms (e.g., codeswitching) of multiple languages in addition to the dominant language within a context including via interlocutors, the media, and across the linguistic landscape (e.g., signage); b) *Multilingualism in Practice*, which includes individual exposure of linguistic diversity as a feature of spoken engagement that one is either directly or indirectly (i.e., ancillary engagement such as overhearing a conversation) involved in; and c) *Linguistic Diversity Promotion*, which is the societal and governmental promotion and encouragement of language variation and use within the context. Both Multilingualism in Context and Multilingualism in Practice scales consist of some items about codeswitching. The former reports on codeswitching about the general language situation within the context, but not about the actual language practice of the participant, while the latter concerns communicative practices of a more personal nature. Moreover, the CILD-Q is part of a larger language profiling questionnaire: The Contextual Linguistic Profile Questionnaire (CLiP-Q), which also assesses demographic information, language history, use and proficiency, as well as socio-economic status (SES). As mentioned, apart from contextual linguistic diversity, most of these factors are usually considered in bilingual studies; however, SES has received less attention.

SES is commonly measured as a proxy of parental and/or self-educational attainment, parental and/or self-occupation, and/or household assets and income, and it has been consistently found to have a large and sustained impact on language development (Brito & Noble, 2014, 2018; Letourneau, Duffett-Leger, Levac, Watson, & Young-Morris, 2013). Unsurprisingly, the trend is that individuals from lower SES backgrounds have poorer linguistic outcomes compared to those from higher SES backgrounds, whereby SES is operationalised as the quality and quantity of the childhood linguistic environment (De Cat,

2020; Fernald, Marchman, & Weisleder, 2013). Given its ubiquitous influence, Wigdorowitz and colleagues (2020) also acknowledged that any measure of language profiling must include SES indicators if it is to provide a comprehensive overview of the population under investigation. Accordingly, factors that characterise the sociolinguistic context and the nature of language input, including SES, cannot be undervalued when the goal is to obtain a best practice perspective able to tease apart the contributions of contextual and individual language experience in order to attain a comprehensive account of linguistic knowledge and influence.

The present study

This study investigates the importance of contextual linguistic diversity when comparing groups from different sociolinguistic contexts, but where English is the *lingua franca*. In this way, we recruited participants from South Africa and England. South Africa is more multilingual than England both in terms of the number of speakers of multiple languages as well as language policy (Constitution of the Republic of South Africa, 1996; Department for Education, 2014; Statistics South Africa, 2012; The British Academy, 2019). Although English is the dominant language in both countries, it is not numerically predominant and may not be the most proficient language of many South Africans, where it is reported as the fourth most common first language (L1 = 9.6%) after Zulu (22.7%), Xhosa (16.0%), and Afrikaans (13.5%) (Statistics South Africa, 2012). While English is the only *de facto* official language of England, reported as the main language for 92% of the population (Office for National Statistics, 2013), 11 languages have attained official status in South Africa (Mesthrie, 2002). In fact, there are more opportunities for active, and crucially passive exposure to multiple languages for South African speakers, whereas the opportunities for

widespread and diverse language use are scarcer and more restricted to particular circumstances and engagements (such as amongst family or religious gatherings) in England.

Both groups were assessed on the CLiP-Q (Wigdorowitz et al., 2020), a language profiling measure designed to capture imperative information about individual and contextual linguistic diversity (regarding both active and passive linguistic experience) as well as additional variables associated with explaining linguistic findings, to understand the relationship between contextual linguistic diversity (CIL-D-Q scales), sociolinguistic context (England vs South Africa), and SES, on the one hand, and contextual linguistic diversity, lingualism status (monolinguals vs bilinguals vs multilinguals), and SES, on the other hand. Finally, we also explored the interplay between sociolinguistic context and lingualism status by analysing codeswitching practice. The specific research questions (RQs) and predictions are as follows:

RQ1. (A) Do people who live in a more multilingual context (South Africa) report greater contextual linguistic diversity than those from a less multilingual context (England)? (B) Do individual differences in SES explain these differences? Given that South Africa is a more multilingual country than England, and that South African speakers have more opportunities for passive linguistic experiences than speakers from England, we predict that South Africans will score higher than participants from England in terms of their contextual linguistic diversity overall, and across all three scales of the CIL-D-Q: Multilingualism in Context, Multilingualism in Practice, and Linguistic Diversity Promotion. We also predict that SES scores will not differ between nationality groups since our sample broadly represents individuals across a range of socio-economic standing in each context. However, given the prevalence of multilingualism in South Africa, we predict that across the socio-economic spectrum, contextual linguistic diversity will be consistently experienced and promoted irrespective of one's SES. In contrast, for the England participants, SES could

influence the uptake of multilingualism in two possible ways. First, those with lower SES will score highly on contextual linguistic diversity, especially if it is the case that lower SES groups use more than one language, as has been reported in many immigrant populations (Fernández Reino, 2019; Lymperopoulou, 2020). As a second alternative, participants in England with high SES will score higher on contextual linguistic diversity because of an awareness that foreign language learning is an index of higher education, global advancement, and cognitive benefit (Hogan-Brun, 2017). This hypothesis would be particularly associated with higher SES families sending their children to independent/private schools, in which foreign language learning is valued and promoted more than in state schools (Tinsley & Doležal, 2018; Tinsley, 2019).

RQ2. (A) Is lingualism status contextually consistent when assessing linguistic diversity in South Africa and England? (B) Do individual differences in SES contribute to any possible differences? Given that contextual linguistic diversity is a phenomenon comprising active and passive linguistic input as derived from the sociolinguistic milieu, we consider it distinct from lingualism status, and therefore predict that self-reported monolinguals from South Africa will score closely to their bilingual and multilingual counterparts on the CILD-Q, who at the same time, are expected to attain close scores (overall and across the three scales), due to the great contextual linguistic diversity of this country. In contrast, because England is a less linguistically diverse country, we predict a larger difference between monolinguals and bilinguals/multilinguals. Regarding the last two groups, it is less clear whether bilinguals and multilinguals will differ from one another, since these participants usually have explicit knowledge of more than one language, and therefore do experience overt linguistic diversity. Therefore, as with South Africans, we do not expect differences on contextual linguistic diversity between bilinguals and multilinguals from

England. Finally, whether individual differences in SES should play a role regarding lingualism status in each country remains a completely exploratory issue.

RQ3. Does codeswitching contribute to the effects of contextual linguistic diversity in South Africa and England across lingualism status groups? As mentioned above, only two of the three CILD-Q scales have items that refer to codeswitching: Multilingualism in Context (MIC) and Multilingualism in Practice (MIP). Recall, MIC reports on the global practice of codeswitching within the context, and in contrast, MIP concerns communicative practices of a more personal nature. Accordingly, the codeswitching analysis focuses on these two scales, distinguishing between non-codeswitching and codeswitching items. Our predictions are as follows. Regarding MIC, we predict that monolingual, bilingual, and multilingual South Africans will score higher on both codeswitching and non-codeswitching items than counterpart lingualism groups from England, given that individuals living in multilingual contexts will be exposed, purely as a consequence of their natural language environment, to more linguistic variation than individuals living in predominantly unilingual contexts. In contrast, for the MIP scale, differences across countries are exclusively expected in the codeswitching items. Specifically, given that this scale relates more directly to one's personal language experience, we predict that South African monolinguals will score higher on codeswitching items than monolinguals from England, since codeswitching is a positive and acceptable linguistic attribute for South Africans across the lingualism spectrum (Mesthrie, 2002), and it has been noted that "monolinguals exposed to linguistically diverse contexts may also switch seamlessly if afforded the opportunity to do so" (Wigdorowitz et al., 2020, p. 5). It is less clear whether bilinguals and multilinguals will differ in their codeswitching practice across South Africa and England since these participants report to have access to more than one language and may engage directly or overhear others engaging in

codeswitching. Despite this, we expect South African bilinguals and multilinguals to score higher on codeswitching items than the same groups from England.

Methods

Participants

The country that the participants reported to have lived in for the longest amount of time guided the classification of the sample into their respective South African and England groups. An initial sample of 353 participants accessed the CLiP-Q but were removed if they had incomplete or anomalous responses, leaving a sample of 269 participants (South Africans = 67.29%, significantly older than England participants). To address the unequal sample size and reduce the age disparity, we computed a propensity score match with age as the matching variable. Matching yielded a final sample of 176 (88 participants across each group; South Africans: $M_{age} = 28.53$, $SD = 7.81$, female = 84.1%; England: $M_{age} = 22.80$, $SD = 6.10$, female = 62.5%)¹. English was reported as the most proficient language of 144 participants and the second most proficient language of 28 participants, with all participants reporting daily English exposure (see Table 1 for English proficiency ratings). Furthermore, English was reported as the primary medium of instruction for the majority of participants across primary ($n = 121$, 70.35%) and secondary ($n = 124$, 70.86%) school settings.

Table 1. Means (and standard deviations) for self-rated English (L1 or L2) proficiency for South African and England participants

English self-ratings	South Africa		England	
	L1 English ($n = 60$)	L2 English ($n = 27$)	L1 English ($n = 84$)	L2 English ($n = 3$)
Speaking	4.88 (0.38)	4.33 (0.73)	4.99 (0.11)	4.50 (1.00)
Understanding	4.95 (0.22)	4.78 (0.42)	4.99 (0.11)	5.00 (0.00)
Reading	4.93 (0.25)	4.78 (0.51)	4.98 (0.15)	5.00 (0.00)

¹ Although nationality groups still differed in age, the matched South African sample had a reduced mean age and resembled the England participants more appropriately.

Writing	4.90 (0.30)	4.44 (0.64)	4.98 (0.15)	4.75 (0.50)
---------	-------------	-------------	-------------	-------------

Note. Scores range from 1 (very low) to 5 (very high).

For those who reported verbal or written proficiency in a language other than English, 43 languages were represented. Our sample was divided into 63 monolinguals (19 South African and 44 England), 55 bilinguals (37 South African and 18 England), and 58 multilinguals² (32 South African and 26 England).

Materials

Contextual Linguistic Profile Questionnaire (CLiP-Q). The CLiP-Q (Wigdorowitz et al., 2020) is a holistic language profiling measure which captures information that has been found to be imperative to language profiling research. The questionnaire includes the following sections:

A. Demographic information. Participants report in which country they have lived in over the majority of their lives, which leads them to answer country-specific demographic questions³ including nationality, country and province/region of birth and current residence, total years lived in reported country, age, gender, and ethnicity.

B. Contextual and Individual Linguistic Diversity Questionnaire (CILD-Q). As mentioned above, the CILD-Q differentiates linguistic diversity within the individual and as a feature of their contextual exposure in relation to the country participants report to have lived in over the majority of their lives (i.e., where they have received the greatest sociolinguistic exposure). Because of its dominance, English is the reference language used to frame the questions. Participants respond, on a Likert scale ranging from 1 (*strongly disagree*) to 5

² These lingualism status descriptors were self-reported.

³ The format and questions are the same for both groups but are worded in a more contextually appropriate manner to ensure clarity and familiarity. For example, level of education question options would match those used in the country.

(*strongly agree*), to 18 items that have been found to reliably measure three scales pertaining to contextual linguistic diversity, including Multilingualism in Context (MIC), Multilingualism in Practice (MIP), and Linguistic Diversity Promotion (LDP; see description above and Table 2 for an overview of the items). Higher mean scores reflect greater general exposure and mixing of different languages from that of English across spoken and written domains (MIC), greater individual and conversational exposure to linguistic diversity (including engagement in code-switching; MIP), and higher governmental and societal encouragement of language diversity (LDP). Item order is randomised across participants.

C. Language history, use, and proficiency. This section is concerned with individual accounts of language history, use, and proficiency and is divided into two components. The first gathers information about general language background including all spoken and written languages, home and most comfortable language, formally and informally learned languages, medium of schooling, exposure to English, and linguicism status self-classification. The second component asks questions about participants' first, second, and third most proficient languages, with repeated questions for each language. Information is obtained about age milestones (e.g., acquisition, writing); years of language use; ability in speaking, understanding, reading, and writing; extent of language use with interlocutors, across activities, and when engaging in personal states (e.g., thinking, doing arithmetic); and degree of cultural association. Lastly, an open-ended question is presented probing participants for any additional language background and usage information they deem to be noteworthy.

D. Socio-economic status (SES). A composite SES score is computed from variables associated with SES, including an index of household assets (home security, computer, paid TV subscription, internet access, car, and domestic worker and/or gardener; 0 = no, 1 = yes, summed), annual household income (value range according to tax brackets; 1 = Less than R195,850/£11,850 to 6 = R1,500,001/£150,001 and above), self, maternal, and paternal level

of education (1 = lower than or up to Grade 11/less than or up to GCSE to 6 = PhD). Scores from each variable were averaged to create a composite SES score ranging from 1 (low SES) to 6 (high SES).

Taken together, the four sections of the CLiP-Q provide a comprehensive linguistic profile of adults situated across various sociolinguistic contexts.

Procedure

Ethical clearance was granted from the authors' institution prior to conducting the study. The CLiP-Q was distributed online across the researchers' networks via Qualtrics and took around 30 minutes to complete. Participation was voluntary and informed consent was requested once participants were informed of the study overview, inclusion criteria, and ethical guidelines.

Data analysis

Linear mixed-effects models (LME) were conducted in *R* (R Core Team, 2020) using the 'lme4' package (Bates et al., 2020), which accounts for both fixed and random effects. LME models can estimate participant and item-level data under one analytic framework, which increases the generalisability of the results (Baayen, Davidson, & Bates, 2008). Participants and items were computed as *random factors* in all models. In addition, *fixed factors* varied according to the RQ. For instance, to address RQ1, Model 1 included nationality (England vs South Africa), CILD-Q scales (MIC vs MIP vs LDP) and SES fixed factors, whereas to address RQ2, Models 2 and 3 included the fixed factors of lingualism status (monolingual vs bilingual vs multilingual), CILD-Q scales (MIC vs MIP vs LDP) and SES, for the South African and England samples separately. SES was considered as a continuous variable, using

centered values (see Schielzeth, 2010, for the rationale of centering). In all cases, the dependent variable was always the mean score (from 1 to 5) obtained from the CILD-Q.

To determine the optimal structure for the random and fixed components of each LME model, the procedure, as outlined by Zuur, Ieno, Walker, Saveliev, and Smith (2009) was followed. We first looked for the best random structure using restricted maximum likelihood (REML), while the full fixed structure (i.e., a three-way interaction in all cases) was retained (Barr, Levy, Scheepers, & Tily, 2013). More specifically, the random structure was tested by running an ANOVA between all possible models containing the various combinations of intercepts and/or slopes, using nationality and lingualism status as random slopes, while keeping the full fixed structure (see also Pérez, Joseph, Bajo, & Nation, 2016, Appendix 2, for the same rationale). The model containing the lowest AIC and BIC values was selected.

Once the best random structure was identified, we then tried to obtain the best fixed structure. To do this, we ran stepwise model comparisons from the most complex model (i.e., three-way interaction) to the simplest model (i.e., main effect), by selecting the significant χ^2 test for the log-likelihood, using maximum likelihood. Third, χ^2 and p values were provided by the ANOVA function of the 'lmerTest' package using Satterthwaite's approximation for denominator degrees of freedom (Kuznetsova, Brockhoff, Christensen, & Jensen, 2020). It was also important to evaluate the effect size of the significant effects, in order to describe the proportion of the total variability attributed to the factor, and in this way, provide an indication of the practical significance of the result. Partial eta squared (η_p^2) was calculated using the 'eta_sq' function of the 'sjstats' package (Lüdtke, 2020). To qualify the two-way interactions, the 'testInteractions' function of the 'phia' package (De Rosario-Martínez, 2015) and 'lsmeans' function of the 'lsmeans' package (Lenth, 2018) was used for post-hoc analyses with Bonferroni correction where necessary. For the three-way interactions, we

divided the data into subsets according to the levels of the CILD-Q scales and fitted adjusted LME for these subsets.

Finally, to address RQ3, we ran independent samples *t-tests* with a Bonferroni correction, comparing lingualism status (monolingual vs bilingual vs multilingual) across nationality (South Africa vs England) in all crossed conditions including either MIC or MIP scales and non-codeswitching and codeswitching items.

Results

The results of the LME analysis are presented in three main sections with a focus on the fixed effects output. Summary details ('lmerTest' package) regarding model fit and random effects of the significant models are provided in Appendix A. We first evaluated whether participants from South Africa and England differed in their contextual linguistic diversity and whether possible differences were explained by SES (RQ1). Secondly, we evaluated whether there were differences of lingualism status within each country and, once more, if individual differences in SES explained their effects (RQ2). Finally, we explored the role of codeswitching in the MIC and MIP scales to investigate differences in lingualism status across South Africa and England (RQ3). Means and standard deviations for item-level data are provided in Table 2, while means and standard error values for CILD-Q scores across nationality groups as a factor of linguistic status are provided in Table 3.

Table 2. Means (and standard deviations) of CILD-Q items across nationality groups

Item	Description	South Africa	England
MIC1	Most people from [country] can communicate in more than one language	4.35 (0.77)	1.84 (0.69)
MIC2 ^R	The local news is only available in English in [country]	4.51 (0.53)	1.85 (1.01)
MIC3	Signs (e.g., public transport) are written in English and at least one other language in [country]	2.88 (1.10)	1.72 (0.99)
MIC4	It is common in [country] to mix words in a	4.16 (0.68)	1.98 (0.91)

MIC5	It is common to mix words from different languages when writing or messaging (e.g., Facebook, WhatsApp)	3.99 (0.84)	2.41 (1.16)
MIC6	People from [country] tend to switch between languages during a conversation	4.16 (0.73)	1.66 (0.74)
MIC7^R	People from [country] tend to speak in the same language throughout a conversation	3.88 (0.86)	1.63 (0.70)
MIP1^R	It is not often that I am exposed to people speaking languages other than English	4.20 (0.92)	3.67 (1.34)
MIP2^R	I can only understand conversations in English	4.05 (0.97)	3.80 (1.49)
MIP3^R	When listening to people speak in a language other than English, I cannot guess what language they are speaking in	3.56 (1.02)	3.98 (1.04)
MIP4	In a conversation I may change from one language to another without even thinking about it	2.97 (1.39)	2.10 (1.27)
MIP5^R	In a conversation I always use the same language throughout	3.13 (1.25)	2.26 (1.16)
MIP6	When speaking with people I know I often use words from different languages	3.49 (1.15)	2.78 (1.24)
MIP7^R	I hardly ever use words from different languages when speaking with people I know	3.57 (1.14)	2.82 (1.38)
LDP1	Speaking multiple languages is encouraged in [country]	4.02 (0.99)	2.64 (1.12)
LDP2^R	Speaking languages other than English is not encouraged in [country]	3.92 (0.97)	2.64 (1.16)
LDP3	The government and people of [country] value speaking more than one language	4.11 (0.82)	2.78 (1.19)
LDP4^R	The government and people of [country] do not value speaking more than one language	4.06 (0.89)	2.78 (1.16)

Note. MIC = Multilingualism in Context. MIP = Multilingualism in Practice. LDP =

Linguistic Diversity Promotion. Codeswitching items in bold. R = reverse-scored item.

[country] = the country the participant reported to have lived in for the longest period.

Table 3. Means (and standard errors) for CILD-Q scores across nationality groups as a factor of lingualism status and CILD-Q scale.

Lingualism status	CILD-Q scale	South Africa	England
Monolingual	MIC	3.95 (0.18)	1.83 (0.19)
	MIP	3.45 (0.18)	3.00 (0.19)
	LDP	3.96 (0.24)	2.67 (0.27)
Bilingual	MIC	3.96 (0.17)	1.84 (0.20)
	MIP	3.57 (0.17)	3.12 (0.20)
	LDP	4.05 (0.22)	2.76 (0.26)
Multilingual	MIC	4.07 (0.17)	1.95 (0.20)
	MIP	3.63 (0.21)	3.19 (0.23)

LDP

3.99 (0.22)

2.70 (0.25)

Note. MIC = Multilingualism in Context; MIP = Multilingualism in Practice; LDP = Linguistic Diversity Promotion.

Contextual linguistic diversity across countries

Regarding differences between countries (Model 1, see Appendix A), our LME model showed a significant main effect of nationality, $F(1, 187.46) = 13.96, p < .001, \eta_p^2 = .07$, where as expected, participants from South Africa ($M = 3.85, SE = .12$) scored higher on the CILD-Q than participants from England ($M = 2.50, SE = .17$). In addition, the two-way interaction of nationality \times scale was marginally significant, $F(2, 184.82) = 2.99, p = .05, \eta_p^2 = .03$, where pairwise comparisons with Bonferroni correction showed that South Africans scored higher than participants from England across two CILD-Q scales: MIC, $t(125.4) = 11.51, p < .001$; and LDP, $t(29.2) = 5.18, p < .001$. No difference was found in the MIP scale ($p = .19$; see Figure 1). To further clarify the meaning of this interaction, a second pairwise comparison analysis was run dividing by nationality. This analysis demonstrated that the three scales did not differ from one another for the South African group ($ps > .55$), but they did for the England group. Specifically, post hoc analyses showed that this group had lower scores in MIC compared to MIP, $t(24.1) = -4.19, p < .01$, but no differences were found between MIC and LDP, nor between MIP and LDP ($ps > .15$).

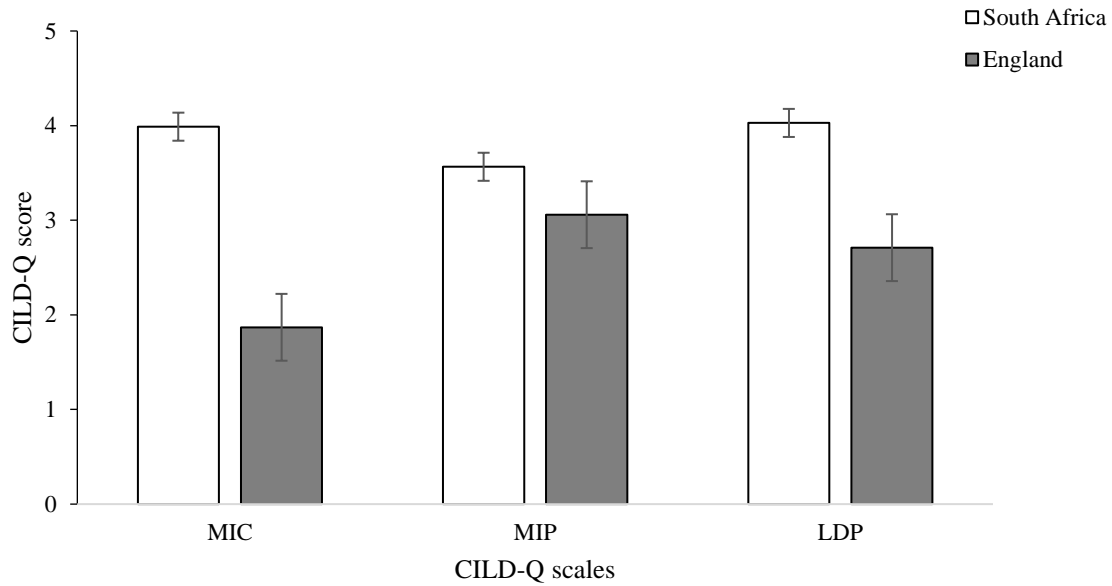


Figure 1. Bar graph with mean and standard error values of CILD-Q scores, representing the two-way interaction between nationality and CILD-Q scales (MIC = Multilingualism in Context; MIP = Multilingualism in Practice; LDP = Linguistic Diversity Promotion).

More importantly, the previous interaction was qualified by a three-way interaction with socio-economic status: nationality \times scale \times SES, $F(2, 174.67) = 3.83, p < .05, \eta_p^2 = .04$. Accordingly, to understand the three-way interaction, a subset was performed by dividing the analysis by scale. This division produced a non-significant interaction of nationality \times SES for both the MIC and MIP scales ($ps > .45$). In contrast, the same interaction was significant in the LDP scale, $F(1, 175.98) = 5.37, p < .05$, where pairwise comparisons with Bonferroni correction showed that SES explained differences in LDP in England, $\chi^2(1) = 6.15, p < .05$, but not in South Africa ($p = .88$). Specifically, England participants with higher SES scored higher in LDP, while those with lower SES scored lower in this scale (see Figure 2). No other effects were significant: scale; SES; nationality \times SES; or scale \times SES (all $ps > .12$).

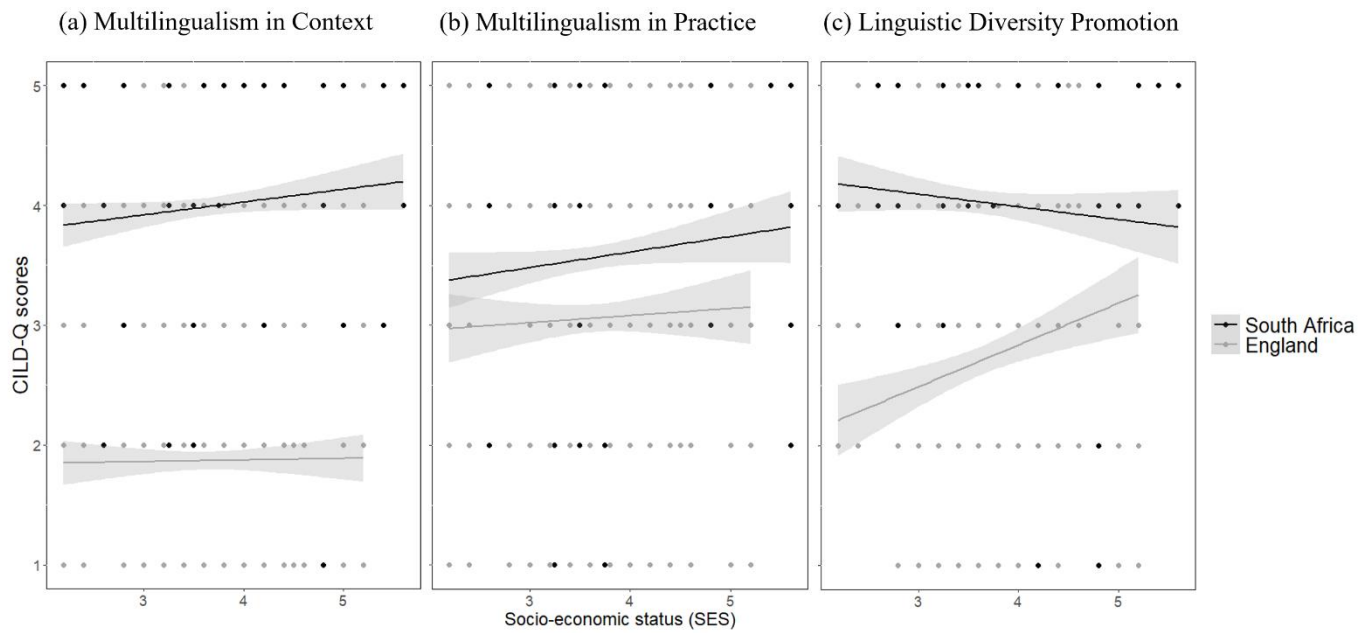


Figure 2. Scatter linear graph with mean CILD-Q scores, representing the interaction between nationality and SES across the three scales: Multilingualism in Context (Figure 2a); Multilingualism in Practice (Figure 2b), and Linguistic Diversity Promotion (Figure 2c).

Altogether, our findings regarding contextual linguistic diversity across countries suggest that compared to England participants, South Africans had greater exposure to linguistic diversity within their contextual confounds and the promotion of linguistic diversity across governmental and societal strata was larger. In addition, contextual linguistic diversity was experienced consistently for South Africans, but the England group varied in how they experienced it. In fact, individual differences in SES showed that contextual linguistic diversity depends on this factor, but only in England which is a less linguistically diverse context. That is, in England, SES appeared to influence whether there was promotion and endorsement of multilingualism, but for South Africans, multilingualism was promoted regardless of one's socio-economic standing.

Contextual linguistic diversity across lingualism status

Taking into account the unequal sample sizes between the two countries across the lingualism conditions⁴, a separate model for each nationality was run to evaluate differences across lingualism status. No significant effects were found in the South African group (Model 2): lingualism status; scale; SES; lingualism status \times scale; lingualism status \times SES; scale \times SES; and lingualism status \times scale \times SES (all $ps > .14$). Regarding differences across lingualism status for England (Model 3, see Appendix A), our LME model manifested a significant effect of scale, $F(2, 23.82) = 10.51$, $p < .001$, $\eta_p^2 = .47$, where pairwise comparisons with Bonferroni correction demonstrated differences between participants across the scales. Specifically, England participants scored significantly lower on MIC in comparison to MIP, $t(25) = 4.12$, $p < .01$, and LDP, $t(23.7) = 2.53$, $p < .05$. In contrast, no difference was found between MIP and LDP ($p = .59$; see England scale scores in Figure 1). No other main or interaction effect was significant: lingualism status; SES; lingualism status \times scale; lingualism status \times SES; scale \times SES; and lingualism status \times scale \times SES (all $ps > .13$).

Taken together, self-reported monolinguals, bilinguals, and multilinguals did not differ from one another in terms of their contextual linguistic diversity in South Africa or England.

Contextual linguistic diversity and codeswitching

Finally, to explore whether some of the differences found across nations and/or lingualism status were a result of items that refer specifically to codeswitching, we ran independent samples t -tests comparing South Africa and England participants across monolinguals, bilinguals, and multilinguals in all crossed conditions including either MIC or MIP scales and non-codeswitching and codeswitching items (see Table 4). A Bonferroni correction for the two types of items separately, set the alpha at .008.

⁴ Our sample comprised 63 monolinguals (19 South African and 44 England), 55 bilinguals (37 South African and 18 England), and 58 multilinguals (32 South African and 26 England).

Table 4. Mean, standard deviation, *p*-value (*t*-test comparison) and *g*-value (Hedges' *g* effect size) for non-codeswitching and codeswitching item clusters of Multilingualism in Context (MIC) and Multilingualism in Practice (MIP) across nationality groups, as a factor of lingualism status.

	Lingualism status	CIL-D-Q scale	South Africa	England	<i>p</i>	<i>g</i>
			<i>M (SD)</i>			
Non-codeswitching	<i>Monolinguals</i>	MIC	3.79 (0.45)	1.80 (0.74)	***< .001	2.95
		MIP	3.79 (0.82)	3.81 (1.03)	.937	-0.02
	<i>Bilinguals</i>	MIC	3.91 (0.58)	1.83 (0.60)	***< .001	3.52
		MIP	4.02 (0.56)	3.76 (1.31)	.432	0.29
	<i>Multilinguals</i>	MIC	3.99 (0.47)	1.78 (0.54)	***< .001	4.34
		MIP	3.93 (0.71)	3.86 (1.08)	.783	0.08
Codeswitching	<i>Monolinguals</i>	MIC	3.93 (0.38)	1.88 (0.73)	***< .001	3.15
		MIP	3.11 (1.10)	2.35 (0.98)	.008	0.74
	<i>Bilinguals</i>	MIC	3.99 (0.59)	1.85 (0.78)	***< .001	3.20
		MIP	3.18 (1.00)	2.76 (1.29)	.193	0.37
	<i>Multilinguals</i>	MIC	4.18 (0.58)	2.04 (0.45)	***< .001	4.02
		MIP	3.52 (1.07)	2.55 (1.02)	***< .001	0.91

Note. Mean scores range from 1-5.

Our results showed that codeswitching played a role depending on the CILD-Q scale. Specifically, South Africans scored higher than England participants on both codeswitching and non-codeswitching items in the MIC scale (all *ps* < .001). Therefore, codeswitching alone did not account for the differences observed in contextual exposure to multilingualism. In contrast, codeswitching explained differences between countries in the MIP scale. That is, there were no significant differences between the two nations across the three lingualism status groups for the non-codeswitching items (all *ps* > .43), whereas codeswitching items revealed differences between lingualism groups across the two countries. More specifically, monolinguals and multilinguals in South Africa scored higher than their England counterparts on the codeswitching items (*ps* = .05 and < .001, respectively, after Bonferroni correction); in contrast the bilingual groups did not manifest significant differences across countries on codeswitching items (*p* = .19). These findings indicate that codeswitching was a feature of

direct communicative practices of South Africans that were self-reported as monolinguals or multilinguals compared to the same populations in England. On the contrary, codeswitching did not seem to be a feature more greatly experienced by South African bilinguals than England bilinguals, suggesting that bilingualism was more similarly defined across the two countries, with equivalent opportunities of codeswitching.

Discussion

A core aim of linguistic research is to, systematically and accurately, describe and quantify language experience. Typically, this is carried out using a language profiling measure where linguistic information of input and usage, age of acquisition and exposure, and proficiency is captured (de Bruin, 2019). Language profiling is largely the first point of data collection in empirical studies aimed at investigating differences across designated language groups. Such information gleaned for these measures plays a vital role in the description, division, and comparison of participants (Luk & Bialystok, 2013; Silva-Corvalán & Treffers-Daller, 2015). Therefore, it is essential that a language profile measure incorporates a broad and valid range of linguistic information that can be used, at minimum, as a baseline for describing an individual's language experience. Accordingly, it was important to assess, using a holistic language profile measure, whether exposure to a predominantly multilingual or unilingual context (where English is the *lingua franca*), affects linguistic knowledge overall, as well as whether self-classification of lingualism status is contextually consistent, taking socio-economic status and codeswitching behaviour into account.

Contextual linguistic diversity across contexts

First, we tried to understand whether people who live in a more multilingual context (South Africa) report greater contextual linguistic diversity than those from a less multilingual

context (England; RQ1A). Unsurprisingly, the results show that the sociolinguistic context of where people are immersed does contribute to their linguistic experience overall even when the prominent language across the contexts is the same. In line with our hypothesis, South Africans reported greater overall contextual linguistic diversity, and more specifically, better multilingual exposure (MIC) and multilingual endorsement (LDP), in comparison to participants from England. Though, it is noteworthy that the differences between nationality groups did not emerge for communicative engagement (MIP; see Figure 1). In this way, we can be confident that contextual linguistic diversity is greatly influenced by exposure to multilingual content within one's environment – whether through the media, on signage, or as a factor of creative and intermingled language practices at the community level – in addition to the value placed on multilingualism from society and the government. The lack of difference between South Africa and England participants on MIP could be due to the fact that this scale relates more to *active* instances of communicative engagement between interlocutors, where an individual may be privy to an exchange either through directly partaking in it or being an overhearer. In this way, for participants who are exposed to additional languages other than English, their exposure here is more explicit, obvious, and may involve unique strategies of communication (such as codeswitching), and as such this is reflected in the smaller mean differences between the nationality groups. Nonetheless, future research should clarify this interpretation.

In addition, South Africans scored similarly across all of the CILD-Q scales, but in England, participants reported greater communicative multilingualism (MIP) in comparison to contextual multilingualism (MIC). This suggests that contextual linguistic diversity is more consistently experienced and pervasive in South Africa, while in England this is not the case. In both countries English attains *societal language dominance* status (Treffers-Daller, 2015) since it is the lingua franca and privileged in terms of use across pedagogy, government,

business, and media. Yet, since South Africa is a unique setting of widespread linguistic fusion, which hosts numerous languages and a diverse array of dialects, we argue that the country's speakers are ubiquitously exposed, both actively and passively, to linguistic diversity in a way distinct from those situated in England. For instance, a South African adult has undoubtedly been surrounded by speakers and content of different languages to that of their L1 (more so, on average, than speakers situated in England), but may not use or have acute knowledge of most or any of these languages. Their knowledge here may be predominantly an awareness of linguistic diversity, but perhaps without actual receptive or productive ability in an additional language, albeit their language skills '*may vary along a continuum from zero to full ability*' (Tsimplici et al., 2020, p. 2). Therefore, the context of language use is an imperative aspect to consider when comparisons are to be drawn about individuals situated in different sociolinguistic settings.

The interplay of multilingual promotion and socio-economic status

We next wanted to address whether SES contributed to the effects of contextual linguistic diversity in South Africa and England (RQ1B). Indeed, we found that the results appear to be dependent on SES, but only in England which is a less linguistically diverse context in both number of speakers of multiple languages as well as policy (Department for Education, 2014; The British Academy, 2019). In England, SES appears to influence whether there is promotion and endorsement of multilingualism, with those of a higher socio-economic standing acknowledging a greater uptake and acceptance of multilingualism than those of a lower socio-economic standing. A possible explanation for this finding can be linked to the notion of what Hogan-Brun (2017) calls *Linguanomics* (the economics of language), whereby multilingualism is viewed as an economic asset and prospect of human capital, as measured both tangibly (e.g., teaching, translation) and intangibly (e.g., culture, identity, human rights).

Our results suggest that those of a high socio-economic standing in England perceive multilingualism as being adequately promoted by the people and country more generally, while those of lower economic advantage perceive multilingualism as being promoted less adequately. It is not the case that the linguistic status of the participants are influencing this finding, since the distribution of those with low SES scores (< 3.5 ; monolinguals = 22, bilinguals = 7, multilinguals = 14) are fairly equal to those of high SES scores (≥ 3.5 ; monolinguals = 22, bilinguals = 11, multilinguals = 12). Rather, the perception of multilingual endorsement is greater for participants with more economic resources, who may therefore have access to such resources through education or other privileged means. Compare, for instance, access to language education in state-funded versus independent schools across England. Although there has been a steady decline in language education across both school sectors, language learning remains more of a pedagogical priority and viable subject option for students attending independent schools (Collen, 2020; Tinsley & Doležal, 2018; Tinsley, 2019). This disparate priority may be one of the factors that accounts for the observed results, since 41% of the sample reported to attend an independent or grammar secondary school, and of those that did attend a state school, 42% had received some secondary education in a medium of instruction other than English. Having the economic and social means to gain experience with language education may, therefore, foster the view that multilingualism is promoted within the country.

Contrastively, for South Africans, multilingualism appears to be promoted by society and the government regardless of one's socio-economic standing. That is, across the low to high socio-economic spectrum, participants equally agree that multilingualism is encouraged. This may be largely influenced by language policies that have been implemented to safeguard multilingualism (e.g., National Language Policy Framework, South African Department of Arts and Culture, 2003), the pedagogical necessity for a multilingual nation, as well as a

pledge of inclusiveness and fair representation, including that of language equity in post-apartheid South Africa (Plüddemann, 2015; Weideman, Read, & du Plessis, 2021). Given that this finding is particularly novel, we encourage further research to investigate these results in more depth to unpack the influence of SES on multilingual promotion.

Contextual linguistic diversity and lingualism status

We next investigated whether lingualism status was contextually consistent when assessing contextual linguistic diversity in South Africa and England separately (RQ2). We found that self-reported monolinguals, bilinguals, and multilinguals did not differ from one another in terms of their contextual linguistic diversity in South Africa or England. That is, South African (England) self-described ‘monolinguals’ scored closely to South African (England) self-described ‘bilinguals’ and ‘multilinguals’ on the CILD-Q, and likewise, the bilinguals and multilinguals scored closely to one another. In line with our hypotheses, a null result of lingualism status on contextual linguistic diversity was present in the South African group, but our findings partially differed from our second hypothesis, where we predicted England monolinguals to score lower than their bilingual and multilingual counterparts given that this is a less linguistically diverse context. In fact, what we did find was a null result for lingualism status on contextual linguistic diversity for England too.

Interestingly, these results suggest that, on the one hand, regardless of the individual lingualism status descriptor one identifies with, their sociolinguistic context is shown to be a consistent contributor to language experience. Contextual linguistic diversity is therefore an imperative variable to consider over-and-above lingualism status, since it is not always possible to show categorical differences across groups that classify themselves under separate linguistic labels. Importantly though, Tsimpli et al. (2020, p. 2) notes that *‘linguistic diversity at the societal level may or may not translate as multilingualism at the individual level’*.

There is thus a distinction between active and passive language experience, but where both contribute holistically to the language repertoire.

On the other hand, this is an illuminating finding, adding evidence to the growing body of literature questioning the appropriateness of a categorical notation of language experience. Recently, *bilingualism* has been disputed as a categorical variable; no definitive experiential threshold suddenly transforms an individual from a monolingual to a bilingual (Luk & Bialystok, 2013; Marian & Hayakawa, 2021; Takahesu Tabori et al., 2018). Rather, language experience is much more complex and multifaceted than the oftentimes reductionist approach it is condensed to. In this sense, our study provides substantial and supplementary evidence supporting this claim, and therefore offers the opportunity to explore similar questions where linguism status is treated on a continuum, rather than categorically.

Clearly, the conceptions of linguism categories are not as bounded as have been previously assumed (for a review see Surrain & Luk, 2019). We have shown that self-descriptions of linguism status in one context can appear similar, thereby providing further evidence illustrating how heterogeneous language groups are, even if people identify themselves under the same linguistic label (also see Beatty-Martínez et al., 2019; Bice & Kroll, 2019; de Bruin, 2019). Accordingly, the more linguistic information we can attain from an individual, the better scope we will have at making some group classifications. We should therefore strive towards a graded approach that positions speakers on a continuum from less multilingual toward more multilingual (Chung-Fat-Yim, Sorge, & Bialystok, 2020; DeLuca, Rothman, Bialystok, & Pliatsikas, 2019; Gullifer et al., 2018; Gullifer & Titone, 2020; Incera & McLennan, 2018; Luk & Bialystok, 2013; Marian & Hayakawa, 2021), specifically with an acknowledgement of their contextual linguistic experience (Wigdorowitz et al., 2020). If we are to continue comparing monolingual and bilingual groups using common categorical methods, then we ought to consider the nuances and complexities that

encompass the language experiences of individuals falling within either of these language groups, and provide a comprehensive and transparent account of as many facets of language information as is possible to attain, particularly highlighting the participants' sociolinguistic context.

Contextual linguistic diversity and codeswitching

Lastly, we aimed to investigate whether codeswitching contributed to the effects of contextual linguistic diversity in South Africa and England across lingualism status groups (RQ3). In line with our hypotheses, we found that codeswitching did not account for the differences across groups in the MIC scale (which addresses more of the general consensus of language use within the context), where all South African lingualism groups scored higher than the same groups from England across items that both contained and did not contain codeswitching. In contrast, codeswitching, as a direct communicative practice, accounted for the higher MIP scores reported by South African multilinguals and monolinguals (this scale refers more specifically to one's personal language practice and engagements). We therefore cannot overlook the fact that there may be a difference in codeswitching practice between South African and England so-called multilinguals, and especially monolinguals. This finding is striking if we assume that monolinguals have access to one language only and codeswitching is generally a proposed bilingual phenomenon (e.g., Beatty-Martínez et al., 2020) that requires access to more than one language.

In addition, there was no difference in MIP codeswitching items for the bilingual groups across the two contexts, nor were there any differences for lingualism groups across the nations on MIP items that did not refer to codeswitching. The fact that codeswitching explained the higher MIP scores reported by multilingual and monolingual South Africans but not for the bilingual group across the contexts may also suggest that bilingualism is a

more concretely defined notion of language knowledge when contrasted with monolingualism and multilingualism, and can be taken to refer specifically to the knowledge of two languages. In contrast, monolingualism and multilingualism are more nuanced notions of language and may be broadly influenced by linguistic (e.g., number of languages spoken, proficiency) and non-linguistic (e.g., community-based norms) factors. Alternatively, these findings further illustrate the problematic categorisation of language knowledge as encompassed within a single linguicism group, as mentioned above.

Essentially the main point of interest is how speakers have perceived themselves in linguistic terms. If it is expected that a typical monolingual or multilingual would not engage in codeswitching because this phenomenon is considered a ‘bilingual’ practice, then it is surprising that South African multilinguals and monolinguals report codeswitching more so than their England counterparts. The sociolinguistic context of where individuals are situated is then clearly an essential contributor to one’s language repertoire. More precisely, individuals situated in multilingual contexts have language experiences that are different from individuals situated in unilingual contexts, even though they may classify themselves under the same linguistic label. Specifically, codeswitching is a common communicative practice in South Africa (Slabbert & Finlayson, 1999; van der Westhuizen & Niesler, 2016), observed across formal and informal settings (Mabule, 2019; Rose & Van Dulm, 2011) and in different communities (McCormick, 2002; Slabbert & Finlayson, 2002), suggesting that it is a widespread and pervasive part of many peoples’ language repertoires. In England, however, it is less common, where it mainly, if at all, occurs within immigrant or heritage communities (e.g., Promprakai, 2018; Wei, 1994). Furthermore, attitudes and community-based norms facilitating (or impeding) codeswitching are also important to consider (Beatty-Martínez, Navarro-Torres, & Dussias, 2020). In multilingual societies, speakers’ intentions to

codeswitch may therefore be driven by pragmatic and interactional opportunities irrespective of one's lingualism status.

Conclusions

While individual reports of language history, use, and proficiency have generally been considered sufficient for language profiling, these variables alone neglect contextual linguistic experience as a factor that contributes to one's overall language repertoire. We have demonstrated that a language profiling measure, such as the CLiP-Q, that captures variables pertinent to language knowledge and experience overall should be a goal towards best practice. When conclusions are to be drawn about the role of factors that contribute to linguistic experience and knowledge, it is imperative that researchers are cognisant of the types of information they have acquired. A full picture cannot be drawn if it lacks an inspection of the sociolinguistic experience and exposure of those being evaluated, given how important context is to one's linguistic repertoire. Fundamentally, we argue that contextual linguistic diversity can influence one's sensitivity to sociolinguistic variation that may be devoid of actual awareness or use of multiple languages, and furthermore, it can shape the status of languages within the socio-cultural context. The nature of contextual linguistic diversity has a profound effect on the linguistic knowledge and use of inhabitants of these contexts. This sociolinguistic factor should be prioritised if research in the language sciences is to make progress.

References

- Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*, 689–698.
- 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.
- 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.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., et al. (2020). *Package "lme4". R package version 1.1-23*. R Core Team. <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- 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.
- Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2019). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication.

- Beatty-Martínez, A. L., Valdés Kroff, J. R., & Dussias, P. E. (2018). From the field to the lab: A converging methods approach to the study of codeswitching. *Languages*, 3(2), 1–19.
- Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language learning. *NeuroReport*, 26(6), 966–971.
- Bice, K., & Kroll, J. F. (2019). English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language*, 196, 104644.
- Brito, N. H., & Noble, K. G. (2014). Socioeconomic status and structural brain development. *Frontiers in Neuroscience*, 8, 1–12.
- Brito, N. H., & Noble, K. G. (2018). The independent and interacting effects of socioeconomic status and dual-language use on brain structure and cognition. *Developmental Science*, 21(e12688).
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249–268.
- Chung-Fat-Yim, A., Sorge, G. B., & Bialystok, E. (2020). Continuous effects of bilingualism and attention on Flanker task performance. *Bilingualism: Language and Cognition*, 1–6.
- Collen, I. (2020). Language Trends 2020: *Language teaching in primary and secondary schools in England*. British Council. Retrieved from https://www.britishcouncil.org/sites/default/files/language_trends_2020_0.pdf
- Constitution of the Republic of South Africa. (1996). *Constitution of the Republic of South Africa, 1996*. Retrieved from <https://www.wipo.int/edocs/lexdocs/laws/en/za/za107en.pdf>
- De Cat, C. (2020). Predicting language proficiency in bilingual children. *Studies in Second Language Acquisition*, 42(2), 279–325.
- 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), 1–13.

- De Rosario-Martínez, H. (2015). *Package “phia”*. R package version 0.2-1. R Core Team.
<https://cran.r-project.org/web/packages/phia/phia.pdf>
- DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7565–7574.
- Department for Education. (2014). *The National Curriculum in England - Framework Document*. Department for Education, available at
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/381344/Master_final_national_curriculum_28_Nov.pdf
- Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248.
- Fernández Reino, M. (2019). *English language use and proficiency of migrants in the UK*. Migration Observatory briefing, COMPAS, University of Oxford, UK. Retrieved from
<https://migrationobservatory.ox.ac.uk/wp-content/uploads/2019/07/Briefing-English-language-use-and-proficiency-1.pdf>
- Freedman, M., Alladi, S., Chertkow, H., Bialystok, E., Craik, F. I., Phillips, N. A., et al. (2014). Delaying onset of dementia: Are two languages enough? *Behavioural Neurology*, *2014*, 1–8.
- Gorter, D. (Ed.). (2006). *Linguistic landscape: A new approach to multilingualism*. Clevedon, England: Multilingual Matters.
- 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.

- Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia* 117, 123–134.
- Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, 23, 283–294.
- Hartshorne, J., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263–277.
- Hogan-Brun, G. (2017). *Linguanomics: What Is the Market Potential of Multilingualism?* London: Bloomsbury Publishing Plc.
- Hopkins, W. G. (2002). A scale of magnitudes for effect sizes: A new view of statistics. Retrieved from <http://www.sportsci.org/resource/stats/effectmag.html>
- Incera, S., & McLennan, C. T. (2018) Bilingualism and age are continuous variables that influence executive function. *Aging, Neuropsychology, and Cognition*, 25, 443–463.
- Kaščelan, D., Prevost, P., Serratrice, L., Tuller, L., Unsworth, S., & De Cat, C. (2020, October 20). A Review of questionnaires quantifying bilingual experience in children: Do we document the same constructs? <https://doi.org/10.31219/osf.io/nry2s>
- Knightly, L. M., Jun, S.-A., Oh, J. S., & Au, T. K.-F. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America*, 114(1), 465–474.
- Kreiner, H., & Degani, T. (2015). Tip-of-the-tongue in a second language: The effects of brief first-language exposure and long-term use. *Cognition*, 137, 106–114.
- 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.

- 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.
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *Package “lmerTest”*. R package version 3.1-2. R Core Team. <https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf>
- Lenth, R. (2018). *Package “lsmeans”*. R package version 2.30-0. R Core Team. <https://cran.r-project.org/web/packages/lsmeans/lsmeans.pdf>
- Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic status and child development: A meta-analysis. *Journal of Emotional and Behavioral Disorders*, *21*(3), 211–224.
- Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second language learning. *Psychological Science*, *20*, 1507–1515.
- Lüdtke, D. (2020). *Package “sjstats”*. R package version 0.18.0. R Core Team. <https://cran.r-project.org/web/packages/sjstats/sjstats.pdf>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*, 605–621.
- Lymperopoulou, K. (2020). Immigration and ethnic diversity in England and Wales examined through an area classification framework. *Journal of International Migration and Integration*, *21*, 829–846.
- 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.

- Marian, V., & Hayakawa, S. (2021). Measuring bilingualism: The quest for a “bilingualism quotient”. *Applied Psycholinguistics* 42, 527–548.
- McCormick, K. (2002). Code-switching, mixing and convergence in Cape Town. In R. Mesthrie (Ed.), *Language in South Africa* (pp. 216-234). Cambridge: Cambridge University Press.
- Mesthrie, R. (Ed.). (2002). *Language in South Africa*. Cambridge: Cambridge University Press.
- Montrul, S. (2015). Dominance and proficiency in early and late bilingualism. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 15–35). Cambridge: Cambridge University Press.
- 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.
- Office for National Statistics. (2013). *Language in England and Wales: 2011*. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/articles/languageinenglandandwales/2013-03-04>
- 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.
- Oh, Y., Todd, S., Beckner, C., Hay, J., King, J., & Needle, J. (2020). Non-Māori-speaking New Zealanders have a Māori proto-lexicon. *Scientific Reports*, 10(1), 22318.
- Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, 22(12), 2728–2744.

- Pérez, A., Joseph, H. S. S. L., Bajo, T., & Nation, K. (2016) Evaluation and revision of inferential comprehension in narrative texts: An eye movement study. *Language, Cognition and Neuroscience*, 31, 549–566.
- 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.
- 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.
- Plüddemann, P. (2015). Unlocking the grid: Language-in-education policy realisation in post-apartheid South Africa. *Language and Education*, 29(3), 186–199.
- Promprakai, N. (2018). *Dispelling myths of first-generation immigrant code-switching: Evidence from Thai marriage migrants in England* (Newcastle University, Newcastle upon Tyne, England). Retrieved from <https://theses.ncl.ac.uk/jspui/bitstream/10443/4239/1/Promprakai%2C%20N.%202018.pdf>
- R Core Team. (2020). *R: The R project for statistical computing*. R Core Team. <https://www.r-project.org/>
- Rose, S., & Van Dulm, O. (2011). Functions of code switching in multilingual classrooms. *Per Linguam*, 22(2), 1–13.
- Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1, 103–113.
- Schmid, M. S. (2013). First language attrition. *Linguistic Approaches to Bilingualism*, 3(1), 94–115.
- Sherkina-Lieber, M., Perez-Leroux, A., & Johns, A. (2011). Grammar without speech production: The case of Labrador Inuttit heritage receptive bilinguals. *Bilingualism: Language and Cognition*, 14(3), 301-317.

- Silva-Corvalán, C., & Treffers-Daller, J. (Eds.). (2015). *Language dominance in bilinguals: Issues of measurement and operationalization*. Cambridge: Cambridge University Press.
- Slabbert, S., & Finlayson, R. (1999). A socio-historical overview of codeswitching studies in the African languages. *South African Journal of African Languages*, 19(1), 60–72.
- Slabbert, S., & Finlayson, R. (2002). Code-switching in South African townships. In R. Mesthrie (Ed.), *Language in South Africa* (pp. 235-257). Cambridge: Cambridge University Press.
- South African Department of Arts and Culture. (2003). *National Language Policy Framework*. Retrieved from http://www.dac.gov.za/sites/default/files/LPD_Language%20Policy%20Framework_English_0.pdf
- Statistics South Africa. (2012). *Census 2011: Census in brief* (Census No. 03-01–41). Statistics South Africa, available at http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf
- Surrain, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, 22(2), 401–415.
- Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, 56, 289–301.
- The British Academy. (2019). *Languages in the UK: A call for action*. The British Academy, available at <https://www.thebritishacademy.ac.uk/sites/default/files/Languages-UK-2019-academies-statement.pdf>

- Tinsley, T. (2019). *Language trends 2019: Language teaching in primary and secondary schools in England survey report*. British Council. Retrieved from <https://www.britishcouncil.org/sites/default/files/language-trends-2019.pdf>
- Tinsley, T., & Doležal, N. (2018). *Language trends 2018: Language teaching in primary and secondary schools in England survey report*. British Council. Retrieved from https://university-council-modern-languages.org/wp-content/uploads/2020/05/language_trends_2018_report.pdf
- Treffers-Daller, J. (2015). Language dominance: The construct, its measurement, and operationalization. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 235–265). Cambridge: Cambridge University Press.
- Tsimpli, I. M., Mukhopadhyay, L., Treffers-Daller, J., Alladi, S., Marinis, T., Panda, M., Balasubramanian, A., & Sinha, P. (2019). Multilingualism and multiliteracy in primary education in India: A discussion of some methodological challenges of an interdisciplinary research project. *Research in Comparative and International Education*, 14(1), 54–76.
- 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.
- van der Westhuizen, E., & Niesler, T. (2016). Automatic speech recognition of English-isiZulu code-switched speech from South African soap operas. *Procedia Computer Science*, 81, 121–127.
- Wei, L. (1994). *Three generations, two languages, one family: Language choice and language shift in a Chinese community in Britain*. Clevedon: Multilingual Matters.

- Weideman, A., Read, J., & du Plessis, T. (Eds.). (2021). *Assessing academic literacy in a multilingual society: Transition and transformation*. Multilingual Matters.
- Wigdorowitz, M., Pérez, A. I., & Tsimpli, I. M. (2020). A holistic measure of contextual and individual linguistic diversity. *International Journal of Multilingualism*.
- Woodcock, R. W., Alvarado, C. G., Rief, M. L., & Schrank, F. A. (2017). Comprehensive Manual. *Woodcock-Muñoz Language Survey III*. Itasca, IL: Houghton Mifflin Harcourt.
- Yu, H. T. (2015). Applying linear mixed effects models with crossed random effects to psycholinguistic data: Multilevel specification and model selection. *The Quantitative Methods for Psychology, 11*, 78–88.
- Zuur, A. F., Ieno, E. I., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York, NY: Springer.

Appendix A. Significant Linear Mixed-Effect (LME) models: Fixed and random effects.

Model 1. LME on the CILD-Q scores including the fixed factors of nationality, scale, and SES

AIC	BIC	logLik	Deviance			
8401.9	8535.3	-4179.0	8357.9			

<i>Random effects</i> ⁺	Variable	Variance	SD			
Participant	Intercept	0.18	0.43			
Item	Intercept	0.16	0.40			
Residual		0.61	0.78			

<i>Fixed effects</i>	β	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>	
Intercept	3.60	0.35	169.00	10.28	***< .001	
England	-1.78	0.48	187.84	-3.72	***< .001	
MIP	-0.51	0.56	182.51	-0.91	.367	
LDP	0.81	0.57	166.92	1.42	.157	
SES	0.11	0.09	174.65	1.26	.211	
England:MIP	1.53	0.77	189.53	1.98	*< .05	
England:LDP	-1.19	0.78	187.37	-1.52	.129	
England:SES	-0.09	0.12	174.65	-0.77	.445	
MIP:SES	0.02	0.14	174.80	0.16	.871	
LDP:SES	-0.21	0.14	174.59	-1.53	.128	
England:MIP:SES	0.02	0.20	174.80	0.12	.907	
England:LDP:SES	0.55	0.20	174.49	2.74	**< .01	

Note. ⁺Random effects based on effect coding, with South Africa and Multilingualism in Context (MIC) scale coded as baseline. MIP = Multilingualism in Practice; LDP = Linguistic Diversity Promotion; SES = socio-economic status.

Model 3. LME on the CILD-Q scores including the fixed factors of lingualism status, scale and SES (England)

AIC	BIC	logLik	Deviance			
4329.7	4388.8	-2153.9	4307.7			

<i>Random effects</i> ⁺	Variable	Variance	SD			
Participant	Intercept	0.26	0.51			
Item	Intercept	0.21	0.46			
Residual		0.64	0.80			

<i>Fixed effects</i>	β	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>	
Intercept	1.87	0.19	21.35	10.03	***< .001	
MIP	1.19	0.27	23.33	4.41	***< .001	
LDP	0.84	0.31	21.71	2.71	*< .05	

Note. ⁺Random effects based on effect coding, with Multilingualism in Context (MIC) scale coded as baseline. MIP = Multilingualism in Practice; LDP = Linguistic Diversity Promotion.