Beyond accuracy

Grose-Hodge, Magdalena; Dabrowska, Ewa; Divjak, Dagmar

DOI:
10.1558/jmbs.23513

Journal of Monolingual
and Bilingual Speech

JMBS (PRINT) ISSN 2631-8407
JMBS (ONLINE) ISSN 2631-8415

Article

Beyond accuracy: Fluency and complexity in the spoken output of bilingual and monolingual preadolescents

Magdalena Grose-Hodge

University of Birmingham, United Kingdom

Ewa Dabrowska

Friedrich Alexander Universitat, Germany; University of Birmingham, United Kingdom

Dagmar Divjak

University of Birmingham, United Kingdom

Abstract

The language of Heritage Speakers (HS), or of early bilinguals of a minority language, is often seen as incomplete or less developed than that of Monolingual Speakers (MS). This study investigates whether 7- to 9-year-old English/Polish HS can be distinguished from MS in terms of linguistic skills when complexity and fluency are focal rather than accuracy. Data from 78 participants shows no significant differences between HS and MS in fluency on an overall measure in either of the languages, although HS produce more fillers and repetitions. On complexity measures, the results for English were similar across groups for Mean Length of T-Unit, but there was a statistically significant difference in Polish, with HS achieving higher values. On a more specific measure of syntactic complexity (Subordination Index), HS achieved higher scores in both languages. There were no significant differences for

Corresponding author

Magdalena Grose-Hodge
Email: magdalena.carlo@gmail.com

Submitted: 26 July, 2022
Accepted: 26 March 2023

https://doi.org/10.1558/jmbs.23513

Lexical Diversity. Corresponding measures were positively correlated across languages, suggesting that the L1 does not impede L2 achievement. Overall, we observed substantial overlap between the groups, with the vast majority of HS falling within the MS norms and the MS falling within the HS norms. This emphasises the need to move away from the deficit approach towards HS.

Keywords: Heritage Language Acquisition (HLA); syntactic complexity; fluency; bilingualism; incomplete acquisition; deficit approach

1. Introduction

With an estimated half of the world's population using more than one language daily, bilingualism is the norm rather than the exception. However, many researchers continue to focus on Monolingual Speakers (MS), and apply monolingual standards to bilingual speakers. Research into differences between monolingual and bilingual performance and processing is not only relatively new, but has also produced conflicting findings, which fuels the existing social ambivalence relating to the acquisition of two languages in childhood. As a result, many parents and teachers worry that speaking a minority language at home or at school may hamper their children's achievement in the dominant language (see Gundarina & Simpson, 2021), despite a large body of more recent research that debunks this common belief (e.g., Papastefanou et al., 2019; Papastergiou & Sanoudaki, 2021). However, studying Heritage Speakers (HS), or early bilinguals of a minority language* (Montrul, 2008, p. 161), is important, not only from the point of educational policymaking, but also for our understanding of the architecture of language cognition. Therefore, this population, originally studied mainly by applied linguists, has now become of interest to theoretical linguists.

1.1. *Reasons to study school-age HS and approaches taken by theoretical linguists*

Relatively little is known about later language development (i.e., during primary school years) in MS, and even less so in HS. For many years, most researchers assumed that children acquire the grammar of their first language fully in their preschool years. Although this assumption has now been shown to be incorrect (e.g., Dabrowska, 2012; Guasti, 2003; Jia & Aaronson, 2003; Pires & Rothman, 2009a), there is relatively little research on later linguistic development. This is

* In this paper, the terms *Heritage Speakers* and *bilinguals* are used interchangeably.

particularly true in the field of Heritage Language Acquisition (HLA). Montrul (2018, p. 534) refers to this period of HLA as 'understudied', emphasising that investigations focusing on children aged 5–17 are necessary to explain what happens between early acquisition and final attainment.

The process of HLA is unlike monolingual acquisition because of the linguistic environment and context it occurs in. While MS grow up exposed to one language only, bilinguals' input is divided between two linguistic codes, with the exposure to the majority language becoming significantly greater with the onset of schooling. Therefore, HS generally become dominant in the majority language, with this dominance shift usually occurring in their early adolescence or even earlier (cf. Papastefanou et al., 2019). The idealised trajectory of HLA (Montrul, 2016, p. 100) thus differs from that attributed to MS.* The motivation for exploring explanations for divergent trajectories and outcomes has led studies to focus on structures in Heritage Language (HL) that differ from monolingual norms.

These differences have been accounted for in various ways. The Incomplete Acquisition Hypothesis (Montrul, 2008; Polinsky, 2006), for example, suggests that certain patterns in the HL, especially those that typically develop later, are not fully acquired by HS. Polinsky (2011) also puts forward attrition on an individual level as an alternative explanation, informed by her study into grammatical knowledge of relativisation in Russian HS in the United States. Additionally, the divergence has been attributed to reduced input and its qualitative properties, as it may be affected by language attrition in caregivers (e.g., Bayram et al., 2019; Montrul & Sánchez-Walker, 2013; Pires & Rothman, 2009b). Lately, explanations focusing on literacy and exposure to formal register have also emerged (Kupisch & Rothman, 2018; Tsimpli, 2014) and the Literacy Enhancement Hypothesis (LEH) was recently proposed by Armstrong and Montrul (2022). According to the LEH, the development of literacy and the experience with written language in school-age children results in more robust representations of morphosyntactic structures, which, in turn, has a positive impact on psycholinguistic processes.

1.2. Assessing HS' and L2 learners' linguistic skills

The focus on identifying HS' linguistic deficits and concentrating on areas of HL that do not converge with the monolingual norm contrasts with current theoretical approaches to assessing linguistic competence in L2 learners

* cf. Rothman (2022) for a different view on HLA trajectories and individual differences; and Dabrowska (2012) for evidence of significant individual differences in monolinguals' linguistic knowledge of key grammatical constructions, indicating trajectories in monolinguals depend on individual differences

(particularly in classroom settings). Researchers influencing classroom practices have emphasised that viewing L2 speakers as failed MS with deficient linguistic systems is problematic (e.g., Cook, 1997), and a shift towards plurilingualism has been observed. Another trend in L2 assessment is a move away from traditional testing, which values grammatical accuracy over other aspects of proficiency, in favour of assessments based on can 'do' approaches, e.g., Communicative Language Testing (Morrow, 2012) and investigating various aspects of proficiency (e.g., the CAF framework: complexity, accuracy and fluency; see Housen & Kuiken, 2009). These theoretical shifts are evident in the Common European Framework of Reference for Languages: Learning, Teaching and Assessment (CEFR; Council of Europe, 2001). A number of researchers (e.g., Benmamoun et al., 2013; Grosjean, 1997; 2008; Kupisch & Rothman, 2018; Pascual y Cabo & Rothman, 2012; Rothman, 2022; Wiese et al, 2022; Zyzik, 2019) have expressed concerns about the tendency towards perceiving HS as failed monolinguals, and, in particular: a) focusing on lower ends of HS competence or overgeneralising from results obtained in the USA; b) focusing on detecting structures that have not been acquired, i.e., centring on HS vulnerabilities and overemphasising accuracy and formal registers, thus possibly magnifying variance; c) benchmarking performance against (often idealised) native speaker standards and using tests developed for and normed on MS; d) taking a unilingual approach (which offers an incomplete picture of early bilingual abilities); and e) overlooking variance in MS.

1.3. Analysing fluency and complexity

Leaving accuracy aside, as it already takes a central position in investigations into HL, this study focuses on the other two dimensions of the CAF Framework, namely *fluency* and *complexity*.

This paper analyses fluency in its narrow sense, defined as 'flow, continuity, automaticity, or smoothness of speech' (Koponen & Riggenbach, 2000, p. 6), which depends on continuous processing between conceptualiser, formulator and articulator (Levelt, 1989). We measure utterance fluency (Segalowitz, 2010) by looking into *speed* (automaticity) and *dysfluencies* (interruptions of the flow between the three systems involved in speech production). These are further divided into *breakdown* and *repair* following Tavakoli and Skehan's (2005) three-dimensional framework.

Bulté and Housen (2012, p. 22) conceptualise complexity as '(1) the number and the nature of the discrete components that the entity consists of, and (2) the number and the nature of the relationships between the constituent components'. This definition informed our choice of complexity measures, which examined both the number and the nature of relationships between the components (see Section 2.1.2).

1.4. The focus of the study and research questions

This study aims to investigate the complexity and fluency of spoken output in school-aged HS and MS. These aspects of HS competency have received much less attention than accuracy. Therefore, we would like to offer a more complete picture of HS abilities and check whether differences between MS and HS are as pronounced as investigations into accuracy in isolation may suggest. This investigation, however, does not see the monolingual group as a control normative standard against which bilingual performance is measured, with the aim to identify HS vulnerabilities. In our main statistical analysis, we examine monolinguals and bilinguals together, taking language experience (bilingualism or monolingualism) as an individual difference, regressed as the main factor with other relevant variables as covariates. We also look beyond accuracy, focusing on the less-emphasised dimensions of competence of the CAF framework, borrowed from the field of L2 acquisition. Furthermore, complexity and fluency measures are derived from spontaneous speech data rather than tests normed for MS. Finally, we take a holistic rather than a unilingual approach by examining both languages the HS use.

Our second aim is to examine the nature of the relationship between fluency and complexity across languages. Some educators (Gundarina & Simpson, 2021) worry that the use of the Heritage Language might be detrimental to the acquisition of the dominant language, and consequently, on immigrant children's academic achievement. Such attitudes were also reported in interviews conducted by us. Some parents were explicitly discouraged from using Polish at home by teachers, and others reported Polish was forbidden on school grounds. If maintaining L1 indeed hampered the development of L2, we would expect a negative relationship between the two languages: i.e., a high level of fluency in one language would be associated with lower levels of fluency in the other language, and analogously for complexity. Therefore, our research questions are:

1) Do monolinguals and bilinguals differ in terms of fluency?
2) Do monolinguals and bilinguals differ in terms of complexity?
3) What is the relationship between fluency and complexity cross-linguistically?

2. Method

We compared complexity and fluency in speech samples of 31 Polish/English HS, 24 English MS and 21 Polish MS aged 8–10 using composite overall measures (Speech Rate and Mean Length of T-Unit), as well as separate measures of fluency (speed, breakdown and repair) and complexity (Lexical Diversity and Subordination Index). Background information about the participants was obtained via parental interviews using the Bilingual Language Experience Calculator (BILEC; Unsworth, 2013). We also administered a receptive grammar test (TROG 2; Bishop, 2003) and its Polish translation to obtain information about the representativeness of our sample. Finally, the bilingual children were also given a Picture Naming Task (PNT) in both languages (see Sections 2.3 and 2.4 for more details).

2.1. Study design

The study design and measures are summarised in Table 1. The dependent variable of interest was Group, i.e., monolingual or bilingual. Age and Socioeconomic Status (SES) were included as covariates, as they are known to correlate with performance of language tasks. All measures used in the study have been shown to increase with age and have been commonly used in the field of language acquisition.

Table 1. Study design summary.

Domain/dimension:	Dependent Measures:		Independent Measure:	Covariates:
COMPLEXITY				
Composite Measure:	*Mean Length of T-Unit*			
Syntax	*Subordination Index*			
Lexis	*Lexical Diversity (VocD/TTR)*	~	**Group**	Age + SES
FLUENCY			mono / bilingual	
Composite Measure:	*Speech Rate*			
Speed	*Articulation Rates*			
Breakdown	*Pauses and Filled Pauses*			
Repair	*Repetitions and Retractions*			

2.1.1. Complexity measures

We used an established measure of complexity, shown to increase throughout childhood and adolescence (Nippold et al., 2005):

- Mean Length of T-Unit in words (MLtUw) – a *T-unit* is defined as a main clause and all subordinate clauses attached to it. The mean length of T-unit is a standard measure of complexity used in the analysis of productive language ability (cf. Hunt 1965; Nippold et al., 2005). It was derived automatically using the MLU function within CLAN

(MacWhinney, 2000), as samples were divided into T-Units during transcription.

Additionally, we chose two more specific complexity measures:

- Subordination Index (SI) – a ratio of the total number of clauses to the number of T-units. Introduced by Loban (1963), who demonstrated that SI increases from kindergarten through to grade 12, this is one of the most common measures of syntactic complexity (Ellis & Barkhuizen, 2005, p. 154). First subordinate clauses and fragments were coded on a comment tier in the transcriptions and counted. Then, the ratio was derived using CLAN (see section on coding for more information on criteria employed for coding subordinate clauses).

Lexical Diversity (LD) 'refers to the range of different words used in a text, with a greater range indicating a higher diversity' (McCarthy & Jarvis, 2010, p. 381). Two measures were used:

- VocD – based on mathematical modelling of the TTR (see below) versus token curve fittings, VocD controls for size by calculating the probability of the frequency of new and repeated lexical items in longer texts (Malvern & Richards, 2002). This measure was derived automatically through the function KidEval within CLAN. However, it is not yet available for Polish, therefore for this language a semi-automatic measure was derived.
- Type to Token Ratio 100 (TTR100) – a bootstrapping approach was taken. Firstly, the function FREQ in CLAN was used on two smaller samples of 100 words. Since we did not have access to a Polish lemmatiser, the list of all word forms generated by CLAN was lemmatised manually and averaged for the two samples.

2.1.2. Fluency measures

Fluency was investigated with a composite overall measure (Speech Rate) and separate measures for speed, breakdown and repair strategies.

- Speech Rate – the mean number of syllables per minute. It was calculated manually over a 2-minute sample.
- Articulation Rate – the mean number of syllables per minute *excluding pauses*. It was derived automatically using the Syllable Nuclei v2 script (de Jong & Wempe, 2009) in PRAAT (Boersma & van Heuven, 2001). The script detects syllable nuclei in running speech by analysing peaks in intensity. Two-minute samples were first edited and cleaned to

reduce background noise and remove any utterances produced by the interlocutor.

- Breakdown Ratio – the mean number of pauses (silent pauses and fillers, e.g., *er, erm*) per minute. The coded symbols on the transcripts have been added up and the resulting number was divided by the length of the entire sample in minutes.

  (1) Silent pauses: **CHI: so the hamster (.) looked (.) at the bee nest .*

  (2) Fillers: **CHI: and the dog is **&-hm** (.) **&-like** (.) pushing the tree.*

- Repair Ratio – the mean number of repairs (i.e., repetitions and retractions) per minute.

  (3) Repetitions: **CHI: <they saw> [/] (.) they saw some trees (.) <with a> [/] (.) with a beehive .*

  (4) Retractions: **CHI: <they tried> [//] they looked over .*

2.2. Participants

Seventy-eight participants (F = 39, M = 39) aged 7;3 to 9;9 (mean = 8;3 [years;month]) were recruited. Thirty-three of them were bilinguals, 24 spoke only English and 21 used exclusively Polish. One of the participants suffered from epilepsy; no other conditions were reported in parental interviews.

2.2.1. Heritage speakers

Thirty-three HS were recruited at a Saturday school in Southampton. Two children were excluded from data analysis; one because her output was unintelligible, and one refused to speak in Polish. Therefore, we analysed data from 31 participants (12 boys, and 19 girls). Twenty-eight participants had two Polish parents; three came from mixed families. Thirty participants were born in the UK, and the remaining one arrived in the UK at the age of 3 months. Polish instruction at the Saturday school consisted of 3–4 hours weekly, focusing on literacy in the HL, as well as Polish culture, history, literature and geography, all taught in Polish. None of the children had any other formal training in their (Geritage) L1.

BILEC interviews revealed that all children were exposed to Polish from birth, and the average age of onset of exposure to English was just over 2 years. The average cumulative exposure was 5.7 years (SD 1.5) for Polish and 2.5 years (SD 1.5) for English. The average current proportion of exposure to Polish, i.e., the percentage of all language input at the time the study was conducted was 50.5% (range 24%–75%, SD 13%). All children were functional bilinguals and the PNT task suggested 17 participants did not have a dominant language, 10 were dominant in English; and four Polish.* The average completed educational levels (computed using parental education levels described following the International Standard Classification of Education Framework, where 5 is equivalent to a university degree) for Carer 1 (usually mother) and Carer 2 (usually father) were 5.23 and 3.6, respectively (below degree-level).

2.2.2. Monolinguals

English MS (16 boys and eight girls) lived in the UK and used only English in their everyday life. Similarly, Polish MS (nine boys and 11 girls) lived in Poland and used exclusively Polish. The groups were similar in age (see Table 2) but while bilinguals and Polish MS had similar SES (mean = 4.4 in each group), the English MS's parents had higher education levels (mean = 5.3; see Table 2). Due to the fact that the study was conducted during the pandemic, all bilinguals and Polish MS were approached at schools, while the recruitment of English MS was carried out on social media, mainly through groups dedicated to educational activities for children in lockdown. This means that the sample is not entirely representative of the general population. Therefore, we added a measure as an additional methodological check, namely the standardised Test for Receptive Grammar (TROG-2) to get a better picture of how (un)representative the sample was. In this test, bilinguals scored similarly to

Table 2. Descriptive statistics for all groups.

	Group	Mean	IQR	Range
Age	Bilinguals	8.28	7.9–8.6	7.3–9.8
	Polish Monolinguals	8.24	7.8–8.5	7.5–9.2
	English Monolinguals	8.18	7.8–8.5	7.3–9.5
SES	Bilinguals	4.4	3.0–5.0	2.5–7.0
	Polish Monolinguals	4.4	3.0–7.0	2.0–7.0
	English Monolinguals	5.4	4.5–6.0	3.0–7.5
TROG Pl	Bilinguals	69.4	67–73	51–77
	Polish Monolinguals	70.1	68–74	62–77
TROG En	Bilinguals	68.7	66–73	56–78
	English Monolinguals	72.8	71–76	60–79

* We used a Picture Naming Task to evaluate this. Pictures were designed to match in both languages in terms of the number of syllables, thematic areas and difficulty. Participants were asked to name all images as quickly as they could without skipping any. A difference in naming latencies of 10 seconds or more between languages was interpreted as an indication of automatisation in lexical access, which informed our decisions on language dominance.

MS in Polish, but their results were slightly lower than those of MS on the English version. The English MS group's score was 15.5, which is between the age equivalent of 9;0 and 10;10, i.e., above their average biological age (8;10). It should also be noted that norms are only given up to age 12, and some participants reached the ceiling.

2.3. Materials

Dependent measures were derived from elicited speech samples. Polinsky (2018) and Montrul (2016) recommend eliciting narratives as the most effective, albeit time- and labour- intensive, measure of proficiency. *The Frog Story*, based on *Frog, Where Are You?* (Mayer, 1969) is the most commonly used tool for sampling spontaneous production. It has frequently been used with MS (Berman & Slobin, 1994), as well as bilinguals (e.g., Treffers-Daller, 2011), including Polish HS (Treffers-Daller & Korybski, 2015). For a description of pictures and the procedure, we refer to the Frog Story Corpora available on CHILDES (MacWhinney, 2000). The most important advantage of this method is that each participant tells almost the same story, so the task demands and the required vocabulary are similar. Another advantage is that the Frog Story allows for cross-linguistic analyses. Finally, this method is suitable for children, as it is enjoyable.

2.4. Procedure

Parental/guardian consent was obtained prior to the study. The children gave their assent and were informed that they could withdraw from the study at any point without giving a reason. Four data collection sessions were conducted (Table 3). During the first session, children were engaged in ice breakers and performed a short PNT in both languages. This was followed by an interview with their parents. In the following sessions, we administered the TROG and collected speech samples. The order in which languages were tested was counterbalanced and the interval between the two sessions was at least 5 days. Children were asked to speak a particular language in each session, and instances of code switching were negligible.

As data collection commenced in 2020, it was affected by the outbreak of the COVID-19 pandemic. Due to school closures and government requirements for social distancing, sessions had to be conducted via Zoom. Three participants were met face-to-face, but the materials were presented on screen to ensure similar conditions. We followed the Berman and Slobin (1994, p. 22) protocol. Transcriptions are available on the OSF repository (https://osf.io/976ej/view_only=eb67ac92d3244a0f9e00773af8f61ac9).

Table 3. Procedure.

Session:	**Tasks:**
Session 1	Introduction, assent, ice breakers, PNT.
Session 2	Interviews with parents (BILEC).
Session 3	Speech samples and TROG 2 – Language 1.
Session 4	Speech samples and TROG 2 – Language 2.

2.5. *Speech sample analysis*

2.5.1. Transcription and coding

Speech samples were transcribed in CLAN (MacWhinney, 2000) and divided into T-units to facilitate the calculation of complexity measures. Transcription and coding were conducted by the first author.

2.5.2. Coding fluency aspects

Breakdowns constituted filled pauses (mainly non-lexical fillers such as *&-erm* but also overused lexical fillers, i.e., *&-like*), while repair strategies consisted of repetitions and retractions. These were marked manually during transcription. Following Schmid and Fagersten (2010), we relied on intuitive criteria for silent pauses, and found that intuitively coded pauses were generally longer than 300 ms; our spot checks in CLAN confirmed this. We did not mark lengths of pauses as it was not feasible given the amount of data we collected. Only pauses and dysfluencies within T-units were coded, as they are more likely to indicate lexical access or grammatical problems. See Section 2.1.2 for examples of fluency coding.

2.5.3. Syntactic complexity

Subordinate clauses were coded manually. Both finite and non-finite clauses were included. However, we excluded non-finite verb complements (e.g., *continue looking, go dancing*), since these are arguably not syntactically more complex than NP complements (e.g., *want to play, go to the shop*). However, infinitival purpose clauses (e.g., *he went out to look for the frog)* were included, as were utterances containing an overt complementiser followed by a clause. We also counted subordinate constructions without a complementiser, provided they were grammatically correct: thus, *the boy told the dog he should be quiet* was counted as subordinate, while *the boy looked worried about the bees were about to attack* was not. Utterances containing only the expression *I think* in main clauses were excluded, as the expression was deemed to function as a parenthetical rather than as a true main clause (Diessel & Tomasello, 2001).

As subordination is a continuum, and some utterances did not follow standard grammatical rules, decisions were not always straightforward. To ensure replicability, rules were compiled and subsequently applied during coding. Details of these are available in the project files on the OSF repository (link above). Any problematic utterances were marked and discussed by a team of linguists before a final decision was made.

3. Results

3.1. Descriptive statistics

Data analysis was performed in R (R Core Team, 2020). Figures 1 and 2 show the distribution of scores on fluency and complexity respectively for monolinguals and bilinguals in both languages. While there were some group differences (in both directions), on the whole, there is a vast amount of overlap between both groups, on all measures.

This impression is confirmed by the data presented in Table 4, which provides information about the number and proportion of HS scoring above, within and below the normal monolingual range (i.e., within two standard deviations of the monolingual mean) as well as MS' scores in relation to bilingual norms.*

3.2. Regression analysis

As the aim of the study was to systematically examine possible differences between MS and HS in terms of fluency and complexity, we fitted an ordinary least squares linear regression model for each of the language measures in each language. The variable of interest was Group (MS or HS). Age and Socioeconomic Status of the parents (obtained by averaging the values for parents) were added as covariates, i.e., variables that account for variance attributable to the factors that are known to influence the outcome but are not relevant to our research question. Thus, the R syntax for each model was *Outcome ~ Group + Age + SES*. This formula was kept constant for all analyses, and insignificant predictors were not removed. Retaining insignificant predictors is more conservative, since removing them tends to lead to an overestimation of the effect of the remaining variables. This approach also allows for direct comparisons of the effect of Group across all models. Multicollinearity was checked by measuring the variance inflation factor using the *car* package in R (Fox & Weisberg, 2019). The homoscedasticity assumption was tested with the Studentized Breusch-Pagan test, run with the *lmtest* package (Zeileis & Hothorn, 2002). To test autocorrelation in residuals, the D-W statistic was calculated, and a Shapiro-Wilk normality test was performed to check

* For normally distributed data, approximately 95% of data points are expected to fall within 2 SDs of the mean, with about 2.5% below and 2.5% above.

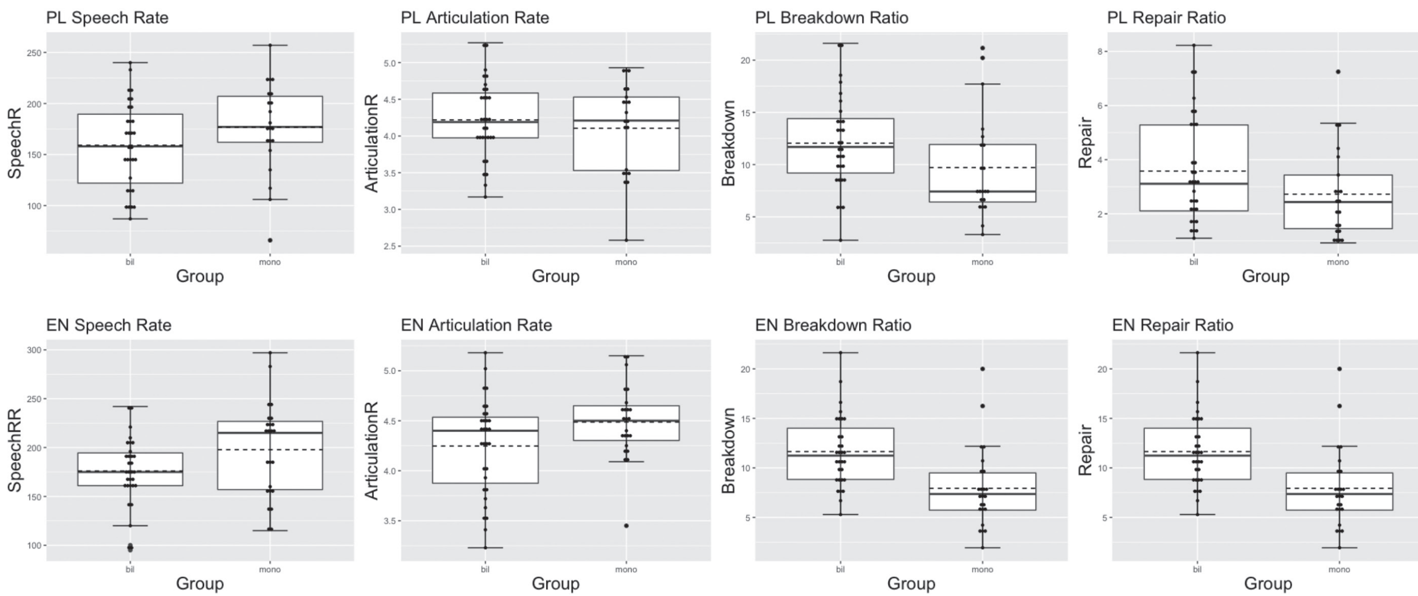

Figure 1. Fluency measures: Results for Polish (top) and English (bottom) with bilinguals (left) and monolinguals (right). Measures from left to right: Speech Rate, Articulation Rate, Breakdown Ratio, Repair Ratio.

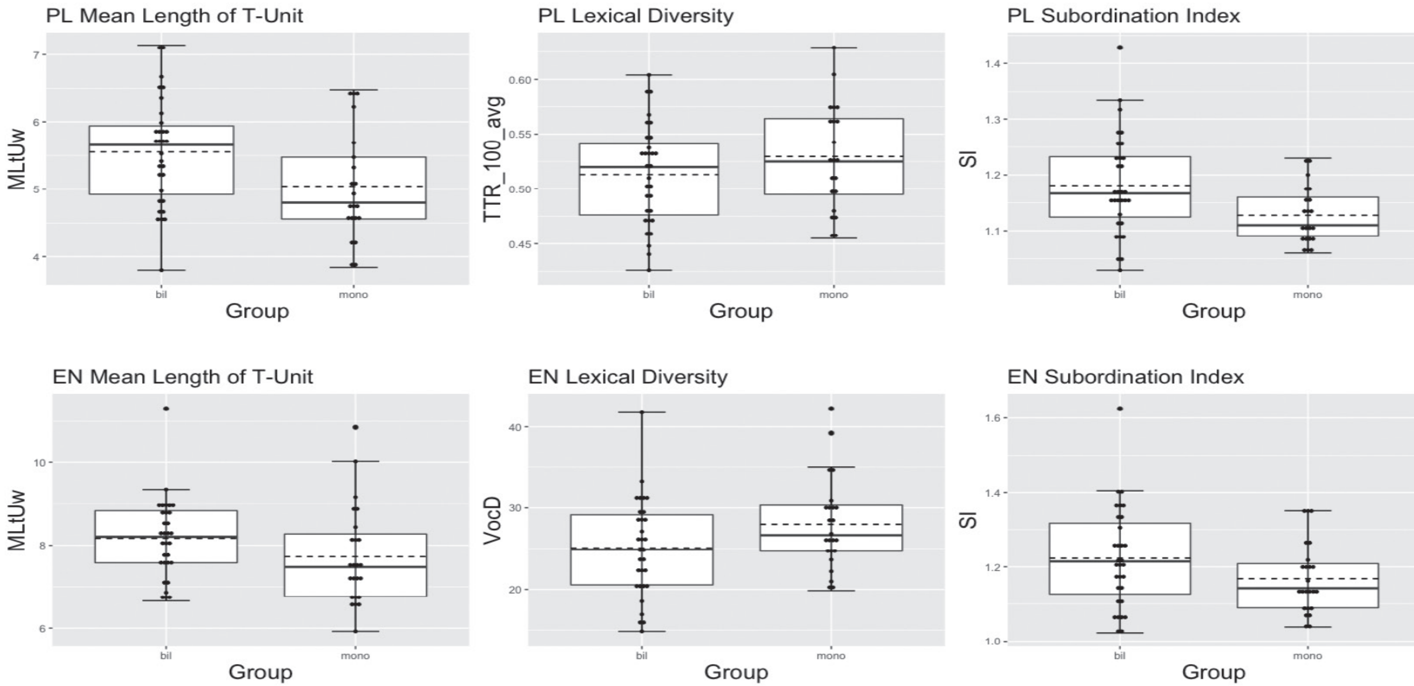

Figure 2. Complexity measures: Results for Polish (top) and English (bottom) with bilinguals (left) and monolinguals (right). Measures from left to right: Mean Length of T-Unit, Lexical Diversity, Subordination Index.

Table 4. The proportion of bilinguals scoring above, within and below the normal monolingual range versus the proportion of monolinguals scoring above, within and below the normal bilingual range.

Bilinguals in monolingual norms						
Polish						
Dimension	**Measure**	**monolingual norms (+/- 2sd)**		**bilinguals below monolingual norms**	**bilinguals within monolingual norms**	**bilinguals above monolingual norms**
Fluency overall	*Speech Rates*	87.21	266.12	1 (3%)	30 (97%)	0
Complexity overall	*Mean Length of T-Unit*	0.81	5.04	0	28 (90%)	3 (10%)
Fluency speed	*Articulation Rates*	2.85	5.36	0	31 (100%)	0
*Fluency breakdown	*Breakdown Ratio*	19.68	–0.26	2 (6%)	29 (94%)	0
*Fluency repair	*Repair Ratio*	6.12	–0.66	4 (13%)	27 (87%)	0
Complexity syntax	*Subordination Index*	1.03	1.23	0	22 (71%)	9 (29)
Complexity lexis	*Lexical Diversity (TTR100)*	0.43	0.63	3 (10%)	30 (97%)	0
	TROG	60.44	79.74	3 (10%)	30 (97%)	0
English						
Fluency overall	*Speech Rates*	100.25	259.59	2 (6%)	29 (94%)	0
Complexity overall	*Mean Length of T-Unit*	5.34	10.1	0	30 (97%)	1 (3%)
Fluency speed	*Articulation Rates*	3.75	5.23	6 (29%)	27 (81%)	0
*Fluency breakdown	*Breakdown Ratio*	16.18	–0.29	3 (10%)	25 (90%)	0
*Fluency repair	*Repair Ratio*	4.69	–0.04	6 (19%)	27 (81%)	0
Complexity syntax	*Subordination Index*	0.98	1.36	0	27 (81%)	6 (19%)
Complexity lexis	*Lexical Diversity (VocD)*	16.84	39.11	3 (10%)	28 (87%)	1 (3%)
	TROG	63.35	82.15	6 (19%)	27 (81%)	0

Monolingual in bilingual norms						
Polish						
Dimension	**Measure**	**bilingual norms (+/- 2sd)**		**monolinguals below bilingual norms**	**monolinguals within bilingual norms**	**monolinguals above bilingual norms**
Fluency overall	*Speech Rates*	73.62	244.25	1 (5%)	19 (90%)	1 (5%)
Complexity overall	*Mean Length of T-Unit*	0.79	5.56	2 (10%)	19 (90%)	0
Fluency speed	*Articulation Rates*	3.17	5.27	1 (5%)	20 (95%)	0
*Fluency breakdown	*Breakdown Ratio*	20.91	3.18	1 (5%)	20 (95%)	0
*Fluency repair	*Repair Ratio*	7.53	–0.37	0	21 (100%)	0
Complexity syntax	*Subordination Index*	1	1.36	0	21 (100%)	0
Complexity lexis	*Lexical Diversity (TTR100)*	0.42	0.6	0	20 (95%)	1 (5%)
	TROG	58.23	80.28	0	21 (100%)	0
English						
Fluency overall	*Speech Rates*	107.96	244.17	0	21 (88%)	3 (12%)
Complexity overall	*Mean Length of T-Unit*	6.24	10.09	1 (4%)	22 (92%)	1 (4%)
Fluency speed	*Articulation Rates*	3.28	5.22	0	24 (100%)	0
*Fluency breakdown	*Breakdown Ratio*	18.92	4.38	1 (4%)	18 (75%)	5 (21%)
*Fluency repair	*Repair Ratio*	7.17	–0.69	0	24 (100%)	0
Complexity syntax	*Subordination Index*	0.95	1.49	0	24 (100%)	0
Complexity lexis	*Lexical Diversity (VocD)*	13.09	37.12	0	22 (92%)	2 (8%)
	TROG	57.68	79.67	0	24 (100%)	0

* Results inverted for repair and breakdown (below norm = more breakdown and repair; above norm = less breakdown and repair)

the distribution of errors. Unless otherwise stated in result tables, all model assumptions were met. The code used for statistical analysis and R output, including model diagnostics, can be accessed in the project files on OSF. The results are presented in Tables 5–8. The tables also include the lmg metric, which is an estimate of the unique contribution of each independent variable to the total variance of the dependent variable, computed using the *relaimpo* package (Grömping, 2006) in R.

3.2.1 Fluency

A regression analysis with the overall composite measure of fluency, Speech Rate (Table 5), shows no statistically significant differences in Group between Polish or English; the only significant variable for this measure is Age.

Table 5. Fluency: regression results for Speech Rates.

		Speech Rates in Polish			
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	–84.09	85.34	–0.985	0.33	
Group	18.810	11.54	1.63	**0.11**	0.042
Age	26.22	9.86	2.66	0.011*	0.108
SES	5.895	3.45	1.71	0.094	0.040
		Speech Rates in English*			
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	–56.189	78	–0.72	0.48	
Group	16.851	11.04	1.527	**0.13**	0.048
Age	23.94	8.72	2.744	0.008**	0.098
SES	7.709	3.87	2	0.051	0.096

* Model diagnostics: model violates the homoscedasticity assumption. An additional generalised linear model confirmed no effect of Group; Pr(>|t|) = 0.133 and a significant effect of Age; Pr(>|t|) = 0.008.

As for individual measures of fluency (Table 6), no significant differences were observed in terms of speed (Articulation Rate) either. Here, Age was again a single significant predictor, but only for English. There is no evidence of statistically significant differences* in terms of breakdown, which again is in line with the result of the analysis of the overall measure of fluency. Finally, in terms of repair, no significant differences were found either.

* The ordinary least squares model fitted for English did show a significant effect of group, but the model diagnostics revealed that the model assumptions were violated. Therefore, a robust linear regression model was run using the function lms in the *MASS* package (Venables & Ripley, 2002), and this did not yield a significant result.

Table 6. Regression results for specific fluency measures.

Articulation Rates in Polish					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	4.675	1.21	3.86	0.0003***	
Group	−0.12	0.164	−0.72	**0.47**	0.011
Age	−0.06	0.14	−0.41	0.67	0.004
SES	0.01	0.05	0.16	0.88	0.001

Articulation Rates in English					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	1.694	0.84	2.03	0.047*	
Group	0.22	0.12	1.87	**0.067**	0.07
Age	0.28	0.09	3.02	0.004**	0.13
SES	0.05	0.04	1.2	0.24	0.03

Breakdown Ratio in Polish					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	−8.028	21.96	−0.37	0.72	
Group	1.34	6.67	−0.2	**0.843**	0.040
Age	2.306	2.08	1.109	0.28	0.100
SES	−0.26	0.49	−0.530	0.6	0.015

Breakdown Ratio in English*					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	39.51	17.97	2.2	0.04*	
Group	−19.25	7.61	−2.53	**0.02***	0.243
Age	−1.89	1.26	−1.502	0.149	0.095
SES	0.450	0.56	0.860	0.400	0.090

* Model should be interpreted with caution as it violates assumptions of the linear model. The Group difference loses significance when robust linear regression is run (rlm function with the MASS package) with t = −2.904.

Repair Ratio in Polish*					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	−1.8447	2.6639	−0.692	0.493	
Group	−0.4405	0.3734	−1.180	**0.245**	0.033
Age	0.5126	0.3046	1.683	0.100	0.061
SES	0.0929	0.1135	0.819	0.418	0.011

* Distribution of errors is not normal. A regression with square root on the outcome variable called with the following syntax: lm (sqrt(repair) ~ Group + Age + SES) confirms no significance of Group; t = 0.099. Other variables are also non-significant.

Table 6 *(continued)*

Repair Ratio in English					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	−1.6	2.177	−0.73	0.47	
Group	−0.19	0.316	−0.6	**0.56**	0.004
Age	0.386	0.240	1.606	0.115	0.044
SES	0.177	0.114	1.561	0.126	0.037

All assumptions of the linear model met but top outliers were removed for bilinguals. In alternative model with all outliers (which does not meet assumptions), the effect of Group was Pr(>|t|) = 0.043.

3.2.2 Complexity

On the overall complexity measure in Polish (Table 7), the Group difference (with bilinguals producing longer utterances) was significant, as was Age. For English, there were no significant factors.

Table 7. Complexity: regression results for Mean Length of T-Unit.

Mean Length of T-Unit in Polish					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	1.67	1.58	1.05	0.3	
Group (mono)	−0.49	0.21	−2.3	**0.026***	0.09
Age	0.49	0.18	2.69	0.010**	0.13
SES	−0.04	0.06	−0.61	0.54	0.01
Mean Length of T-Unit in English					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	6.246	2.007	3.11	0.003**	
Group	−0.417	0.29	−1.44	**0.156**	0.034
Age	0.172	0.23	0.768	0.45	0.010
SES	0.09	0.1	0.87	0.387	0.007

For specific measures of complexity (Table 8), results revealed a similar pattern for both languages. While there were no differences in Lexical Diversity, in terms of syntactic complexity, the effect of Group was significant confirming that bilinguals' output was more complex as they produced more subordination.

3.2.3 Cross-linguistic correlations

Finally, to examine the relationships between the linguistic measures in both languages in the bilingual participants, we computed Pearson correlations between the Polish and English scores. As shown in Table 15, all the correlations were positive and statistically significant.

Table 8. Complexity: regression results for separate measures of fluency.

Subordination Index in Polish					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	1.13	0.164	6.88	1.11e-08***	
Group	−0.0053	0.02	−2.4	**0.020***	0.108
Age	0.006	0.02	0.31	0.76	0.002
SES	0.002	0.01	0.25	0.8	0.001
Subordination Index in English					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	1.13	0.237	4.77	1.56e-05***	
Group	−0.0762	0.03	−2.27	**0.027***	0.070
Age	0	0.03	−0.05	0.96	0.001
SES	0.02	0.01	1.98	0.053 .	0.051
Lexical Diversity (TTR_100) in Polish					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	0.389	0.1	3.99	0.0002***	
Group	0.017	0.01	1.3	**0.2**	0.031
Age	0.02	0.01	1.43	0.16	0.040
SES	0	0	−0.55	0.58	0.008
Lexical Diversity (VocD) in English					
Variable	**Parameter estimate**	**Standard error**	***t* value**	**Pr(>|t|)**	**lmg**
Intercept	31.01	11.7	2.65	0.011**	
Group	2.07	1.66	1.25	**0.22**	0.042
Age	−1.11	1.31	−0.85	0.4	0.020
SES	0.75	0.580	1.29	0.2	0.049

Table 9. Cross-linguistic correlations between measures in bilinguals.

Measures	***r***	***t***	***df***	***p***
Fluency overall *Speech Rate*	0.36	2.10	29	0.04
Complexity overall *Mean Length of T-Unit*	0.67	4.83	29	4.067e-05
Fluency Speed	0.52	3.31	29	0.002
Fluency Breakdown	0.42	2.47	29	0.02
Fluency Repair	0.80	7.28	29	5.119e-08
Complexity syntax *Subordination Index*	0.59	3.97	29	0.0004
Complexity vocabulary *Lexical Diversity*	0.44	2.61	29	0.01

4. Discussion

When interpreting the above results, it should be taken into consideration that our bilinguals had considerably less input than our monolingual participants. Recall that the HS' cumulative exposure to English was on average just over 2 years (mean = 2.5 years, median = 1.9 years), compared to over 8 years in MS (mean = 8.2, median = 8.2), and that MS were well above norms for their age.

4.1. Fluency

Our results suggest no significant differences between the groups relating to speech rates, speed of articulation or breakdown in Polish or English despite the differences in the amount of input and SES.

The values for all dysfluencies were correlated positively with age in all groups, i.e., older children produced more dysfluencies (r = 0.20 for English MS and 0.17 for Polish MS; for bilinguals, the corresponding figures were 0.32 for English and 0.15 for Polish), which shows that a higher number of dysfluencies does not translate into delayed development.

Further analysis of dysfluencies indicates that the largest differences between the groups were observed in the use of repetitions (repair), with bilinguals producing almost twice as many (2.12 per minute in Polish and 1.95 in English) as MS (1.3 per minute in Polish and 0.95 in English), and fillers (breakdown), with bilinguals again producing more (3.2 per minute in Polish and 1.8 in English) than MS (2.0 per minute in Polish and 0.8 in English). For unfilled pauses and retractions, the differences were less clear. Corpus based studies looking into the relationship between filled pauses and complexity have shown that fillers (e.g., *erm*, *er*) and repetitions often precede longer and more complex constituents (e.g., Clark & Wasow, 1998; Cook et al., 2016; Watanabe et al., 2005). Therefore, it is possible that bilinguals produce more of this type of dysfluencies, as planning utterances that are complex syntactically entails a heavier cognitive load (also see Section 4.2). This can be seen in the following examples:

(5) **CHI: <and they promised that>* ***[/]*** *(.) and they promised <that> [//] (.) the baby frog's parents (.) that they will take very care of him.*
(6) **CHI: <and the boy> [/] and the boy (.) he was looking in a hole to see if the frog was there.*
(7) **CHI: the boy decided to* ***&-erm*** *(.) escape the owl by climbing up on a rock.*
(8) **CHI: they crawled* ***&-erm*** *[/] (.) they crawled out of the* ***&-erm*** *[/] (.) out of the lake and found that they hear something behind the log.*

4.2. Complexity

Perhaps the most interesting result in terms of complexity is that bilinguals achieved higher scores on measures of complexity *in both languages*. To our knowledge, this is the first study to report such findings. Preliminary qualitative analyses of the samples suggest bilinguals' boldness and creativity and at the same time less adherence to linguistic conventions. See examples below:

(9) **CHI: <they crawled> &-erm [/] (.) they crawled <out of the> &-erm [/] (.) out of the lake and* ***found that they hear something behind the log****.*

(10) **CHI: and the dog (.)&-uh looks shocked about why* ***there are so many wasps****.*

(11) **CHI: the dog climbs on top of the boy's head* ***because he (.) looks like he doesn't (.) like getting wet****.*

(12) **CHI: <the dog searched in> [/] the dog searched in the jar (.)* ***including putting his head into it****.*

The complexity and creativity of their language, coupled with lower idiomaticity, could be an indirect consequence of lower exposure to each language or, conversely, exposure to more varied or richer input containing constructions in two languages.

The first explanation, which looks at the input in each language separately, could be found in the role of entrenchment, and in the way in which children slowly eliminate ungrammatical structures from their linguistic repertoire without negative evidence from carers and teachers. Brooks and Tomasello (1999) tested monolingual children aged 3 to 8 and observed that low frequency verbs are more likely to be used in a novel way (e.g., *the magician vanished the rabbit*) than high frequency ones (*the magician disappeared the rabbit*). Ambridge et al. (2015) asked English-speaking children and adults to rate sentences that included overgeneralisation errors (e.g., *daddy giggled the baby*). They found that the higher the overall frequency of the verb in an incorrect structure, the less likely the participants were to accept it as accurate. These studies provide evidence that frequency and entrenchment play a great role in retreat from error and in inhibiting the production of novel or non-conventional structures. It is possible that, since our bilinguals receive less input in either language than MS, their language is less conventional and therefore appears as more innovative and creative. In other words, since they had less experience with the languages, linguistic structures were less entrenched in their minds, so they were more adventurous in the way they used them. Bilinguals in this sample had the experience of a 2-year-old with the English language but they had cognitive abilities of an 8-year-old, which allowed them to produce more complex structures that reflect the stage of the cognitive development they are at without the conventional constraints that more exposure and therefore entrenchment would result in.

Goldberg (2019) provides a slightly different account of the emergence of linguistic skills that are creative yet constrained by convention, which is not based on frequency. She explains why structures such as *explain this to me* are preferred over *explain me this* in terms of coverage (mutual similarity between different instantiations of a construction) and statistical pre-emption (competition between constructions). Abstract exemplars that represent linguistic knowledge (form-function pairings) cluster together and compete in the hyper-dimensional conceptual space where they are stored. When activated simultaneously, these clusters compete, and the more conventional alternative eventually pre-empts the alternative. Building on this, we observe that in bilingual minds, there are more alternatives available to express the same message, and therefore, the process of pre-emption may work differently. We know that transfer phenomena are ubiquitous in bilinguals' speech, suggesting that they rely less on pre-emption and therefore are less bound by conventional constraints and more ready to explore alternative constructions.

4.3. Cross-linguistic correlations

Taking into consideration the concerns surrounding the use of minority languages at school and home (see Section 1), we tested the hypothesis that corresponding measures would correspond negatively, i.e., high competence in one language would interfere with competence in L2. As we have seen, the cross-linguistic correlations for all measures were actually positive. This finding, coupled with the fact that bilinguals attained similar levels of competence compared to monolinguals, demonstrates that a high level of L1 is not achieved at the expense of or to the detriment of L2 development. Therefore, the use of L1 should not be a cause for concern for educators and parents, as there is no reason to believe that it has a detrimental effect on the acquisition of the L2, and, consequently, on academic performance. While lower achievement in children from immigrant families has been observed in some contexts, any gap is likely to be connected to non-linguistic factors, such as socioeconomic status, public perceptions of immigration (e.g., Finch et al., 2021) or school-parent relationships. It could also be exaggerated by including results of children who have only just arrived and are tested in a language they do not yet speak. Crosnoe and Turley (2011) found that when SES is controlled for, the disparities disappear, and in fact, immigrant children perform relatively better. Moreover, many studies report an immigrant advantage (e.g., Hao & Woo, 2012). The positive correlations are consistent with results reported in other studies (e.g., Papastergiou & Sanoudaki, 2021; Pham, 2016; but cf. Simon-Cereijido & Gutierez-Clellen, 2009), and the fact that many bilinguals perform as well if not better than their monolingual peers, despite lower exposure, strongly suggests a positive transfer of L1 skills to L2. Demie and Lewis (2018), who have analysed EAL student achievement over years, initially looking at data from a Local Education Authority in inner London, and subsequently national statistics, observed a trend: while EAL student achievement at an early stage of English acquisition is below the expected standard, fully fluent EAL learners outperform their monolingual peers in national tests at the end of primary (e.g., Demie & Strand, 2005) and secondary school (e.g., Demie, 2017; Strand & Demie, 2006) in all subjects. This tendency has also been reported in a recent government document (Department of Education, 2020), which analysed the results of formal tests implemented nationwide at EYFS, KS1, KS2 and KS4 in the 2017/2018 academic year. The use of minority languages in state schools could, therefore, bridge the achievement gap at the early stages of English acquisition, and accelerate, not impede, the development of English, as students are more likely to learn L2 if the concepts discussed in the language are accessible to them.

4.4. Towards a more balanced approach to studying HS' abilities

It is therefore imperative that public perceptions are informed by studies that employ methods and frameworks that embrace diversity and avoid pathologising minorities (see Section 1.2). Moreover, a commitment to equitable and objective approaches to studying HS would not only inform educational policies and exert a positive influence on attitudes to bilingualism, but could also enhance our understanding of language and cognition. First, avoiding normative monolingual comparativity, which frames MS as ideal speaker-hearers and owners of the language, may produce a more complete or accurate picture of linguistic skills of both HS and MS (see Wiese et al., 2022 for evidence that descriptions of heritage grammars found in literature misattribute some non-canonical patterns to bilinguals although they are equally present in monolinguals). This includes avoiding measures that disadvantage bilinguals. It is worth noticing here that the only test on which bilinguals' scores were considerably lower in this study was the TROG-2 (receptive grammar), i.e., the only instrument developed for and normed on MS that we employed. Furthermore, this contrasted with the results obtained on productive grammar, which is surprising, as the population is believed to have stronger receptive skills. Finally, venturing outside testing HS' deficiencies would allow studies to more fully explore phenomena that have not yet received full attention due to the methods that have been used in the field, e.g., variation within the HS population (Rothman et al., 2022), language variation and change (Wiese et al., 2022), linguistic innovation (Zyzik, 2019) or the process of negotiating linguistic conventions in speakers' minds.

5. Conclusion

There is a vast overlap between the MS and HS populations, and the language of HS is syntactically highly complex. Additionally, the positive relationship between performance in both languages, coupled with the fact that HS could be distinguished from MS only on measures in which they outperformed MS, proves concerns that maintaining a heritage language would hamper the acquisition of L2 are ungrounded . While poor academic achievement may be linked to various factors, in case of bilingual students, it is often hastily attributed to their linguistic background. In reality, limiting HS' opportunities to maintain their family language is likely to have a detrimental effect not only on their (L1 and L2) linguistic skills and academic achievement, but also on their emotional development, as it could affect their relationships with extended family and interfere with their heritage identity. It is, therefore, imperative that public perceptions and advice given by educators are based on reliable and unbiased academic research.

The results of our study also imply that exploring the whole range of HS' linguistic skills, instead of focusing on detecting patterns that have not (yet) been acquired, provides a different insight into their linguistic abilities, and highlights novel areas worth investigating, leading to a better understanding of the population and the nature of the process of language acquisition.

References

Ambridge, B., Bidgood, A., Twomey, K. E., Pine, J. M., Rowland, C. F., & Freudenthal, D. (2015). Preemption versus entrenchment: Towards a construction-general solution to the problem of the retreat from verb argument structure overgeneralization. *PLoS ONE*, *10*(4), e0123723. https://doi.org/10.1371/journal.pone.0123723

Armstrong, A., & Montrul, S. (2022). The impact of schooling on bilingual children's language development. Paper presented at the 4th International Symposium on Monolingual and Bilingual Speech, University of Louisiana at Lafayette.

Bayram, F., Rothman, J., Iverson, M., Kupisch, T., Miller, D., Puig-Mayenco, E., & Westergaard, M. (2019). Differences in use without deficiencies in competence: Passives in the Turkish and German of Turkish heritage speakers in Germany. *International Journal of Bilingual Education and Bilingualism*, *22*(8), 919–939.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Defining an "ideal" heritage speaker: Theoretical and methodological challenges: Reply to peer commentaries. *Theoretical Linguistics*, *39*, 259–294. https://doi.org/10.1515/tl-2013-0018

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study.* Lawrence Erlbaum.

Bishop, D. (2003). *Test for reception of grammar, version 2 (TROG-2)*. Pearson Assessment.

Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with Praat. *Glot International*, *5*(9–10), 341–347.

Brooks, P., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, *75*, 720–738.

Bulté B., & A. Housen. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37,* 201–242. https://doi.org/10.1006/cogp.1998.0693

Cook, M., Smith, J., & Lalljee, M. G. (2016). Filled pauses and syntactic complexity. *Language and Speech*, *17*(1), 11–16. https://doi.org/10.1177/002383097401700102.

Cook, V. (1997). Monolingual bias in second language acquisition research. *Revista Canaria de Estudios Ingleses*, *34*(1), 35–50.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Crosnoe, R., & Turley, R. N. L. (2011). K-12 educational outcomes of immigrant youth. *The Future of Children*, *21*(1), 129–152. https://doi.org/10.1353/foc.2011.0008

Dabrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253.

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Demie, F. (2017). English as an additional language and attainment in primary schools in England. *Journal of Multilingual and Multicultural Development, 39* (3), 210-223. https://doi.org/10.1080/01434632.2017.1348508

Demie, F., & Lewis, K. (2018). Raising achievement of English as additional language pupils in schools: Implications for policy and practice. *Educational Review*, *70*(4), 427–446.

Demie, F., & Strand, S. (2005). English language acquisition and educational attainment at the end of primary school. *Educational Studies*, *31*(3), 275–291. http://doi.org/10.1080/03055690500236613

Department of Education (2020). *English proficiency of pupils with English as an additional language.* Department of Education, UK at www.gov.uk/government/publications.

Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics*, *12*(2), 97–141. https://doi.org/10.1515/cogl.12.2.97

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University Press.

Finch, H., Hernández Finch, M. E., & Avery, B. (2021) The impact of national and school contextual factors on the academic performance of immigrant students. *Frontiers in Education*, *6*(793790). https://doi.org/10.3389/feduc.2021.793790

Fox J., & Weisberg S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions.* Princeton University Press.

Groemping, U. (2006). Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software, 17*(1), 1–27. https://doi.org/10.18637/jss.v017.i01Grosjean, F. (1997). The bilingual individual. *Interpreting*, *2*, 163–187. https://doi.org/10.1075/intp.2.1-2.07gro

Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.

Guasti, M. T. (2003). Relative clause formation in Romance childs production. *Probus*, *15*(1), 47–89. https://doi.org/10.1515/prbs.2003.005

Gundarina, O., and Simpson, J. (2021). A monolingual approach in an English primary school: Practices and implications. *Language and Education*, *36*(6), 523–543. https://doi.org/10.1080/09500782.2021.1945084.

Hao, L., & Woo, H. S. (2012). Distinct trajectories in the transition to adulthood: Are children of immigrants advantaged? *Child Development*, *83*(5), 1623–1639. https://doi.org/10.1111/j.1467-8624.2012.01798.x

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461–473.

Hunt, K. (1965). *Grammatical structures written at three grade levels*. Urbana.

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*(1), 131–161. https://doi.org/10.1017/S0142716403000079

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). University of Michigan Press.

Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, *22*(5), 564–582. https://doi.org/10.1177/1367006916654355

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Loban, W. (1963) *The language of elementary school children*. NCTE Research Report No. 1. National Council of Teachers of English.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85–104. https://doi.org/ 10.1191/0265532202lt221oa

Mayer, M. (1969). *Frog, where are you?* Penguin Young Readers Group.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor.* John Benjamins.

Montrul, S. (2016). *The acquisition of heritage languages.* Cambridge University Press.

Montrul, S. (2018). Heritage language development: Connecting the dots. *International Journal of Bilingualism*, *22*(5), 530–546. https://doi.org/10.1177/1367006916654368

Montrul, S., & Sánchez-Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition*, *20*(2), 109–132. https://doi.org/10.1080/10489223.2013.766741

Morrow, K. (2012). Communicative language testing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 140–146). Cambridge University Press.

Nippold, M. A., Hesketh, M. J., Duthie, J. K., & Mansfield, T.C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1048–1064. https://doi.org/10.1044/1092-4388(2005/073)

Papastefanou, T., Powell, D., & Marinis, T. (2019). Language and decoding skills in GreekEnglish primary school bilingual children: Effects of language dominance, contextual factors and cross-language relationships between the heritage and the majority language. *Frontiers in Communication*, *4*, 65. https://doi.org/10.3389/fcomm.2019.00065

Papastergiou, A., & Sanoudaki, E. (2021). Language skills in Greek-English bilingual children attending Greek supplementary schools in England. *International Journal of Bilingual Education and Bilingualism*, *25*(8), 2834–2852. https://doi.org/10.1080/13670050.2021.1980496

Pascual y Cabo, D., & Rothman, J. (2012). The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics*, *33*, 450–455.

Pham, G. (2016). Pathways for learning two languages: Lexical and grammatical associations within and across languages in sequential bilingual children. *Bilingualism*, *19*(5), 928–938. https://doi.org/10.1017/S1366728915000899

Pires, A., & Rothman, J. (2009a). *Competence divergence among monolingual speakers: The role of late triggers*. Unpublished manuscript, University of Michigan and University of Iowa.

Pires, A., & Rothman, J. (2009b). Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism*, *13*(2), 211–238.

Polinsky M. (2006). Incomplete acquisition: American Russian. *Journal of Slavic Linguistics*, *14*, 191–262.

Polinsky, M. (2011). Reanalysis in adult heritage language: A case for attrition. *Studies in Second Language Acquisition*, *33*, 305–328. https://doi.org/10.1017/S027226311000077X

Polinsky, M. (2018). *Heritage languages and their speakers.* Cambridge University Press.

R Core Team. (2020). *R: A language and environment for statistical computing. Manual.* R Foundation for Statistical Computing.

Rothman, J. (2022) *Individual differences, context and support in heritage (minority) language bilingualism: What European studies suggest.* University of Florida Press.

Rothman, J., Bayram, F., DeLuca, V., Di Pisa, G., Duñabeitia, J., Gharibi, K., Hao, J., Kolb, N., Kubota, M., Kupisch, T., Laméris, T., Luque, A., van Osch, B., Pereira Soares, S. M., Prystauka, Y., Tat, D., Tomić, A., Voits, T., & Wulff, S. (2022). Monolingual comparative normativity in bilingualism research is out of "control": Arguments and alternatives. *Applied Psycholinguistics*, *44*(3), 1–14. https://doi.org/10.1017/S0142716422000315

Schmid, M. S., & Fägersten, K. B. (2010). Disfluency markers in L1 attrition. *Language Learning*, *60*(4), 753–791.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexicon and grammar. *Applied Psycholinguistics*, *30*, 315–337.

Strand, S., & Demie, F.. (2006). English language acquisition and educational attainment at the end of secondary school, *Educational Studies*, *32*(2), 215–231. http://doi.org/10.1080/03055690600631119

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance* (pp. 239–273). John Benjamins. https://doi.org/10.1075/lllt.11.15tav

Treffers-Daller, J. (2011). Operationalizing and measuring language dominance. *International Journal of Bilingualism*, *15*(2), 147–163. https://doi.org/ 10.1177/1367006910381186

Treffers-Daller, J., & Korybski, T. (2015). Using lexical diversity measures to operationalize language dominance in bilinguals. In C. S. Corvalan, & J. Treffers-Daller (Eds.), *Language dominance in bilinguals*. Cambridge University Press.

Tsimpli, I. M. (2014). Early, late or very late?: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism*, *4*, 283–313. https://doi.org/10.1075/lab.4.3.01tsi

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*(1), 86–110. https://doi.org/10.1017/S1366728912000284

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2005). Filled pauses as cues to the complexity of following phrases. In *Proceedings of the 9th European Conference on Speech Communication and Technology* (pp. 37–40). ISCA. https://doi.org/10.21437/interspeech.2005-33

Wiese, H., Alexiadou, A., Allen, S., Bunk, O., Gagarina, N., Iefremenko, K., ... & Zuban, Y. (2022). Heritage speakers as part of the native language continuum. *Frontiers in Psychology*, *12*, 717973.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10.

Zyzik, E. (2019). Incomplete acquisition from a usage-based perspective: A response to Dominguez, Hicks and Slabakova. *Studies in Second Language Acquisition*, *41*(2), 279–282. https://doi.org/10.1017/S0272263119000330