The Persistence and Functional Impact of English Language Difficulties Experienced by Children Learning English as an Additional Language and Monolingual Peers

Katie E. Whiteside[a] and Courtenay Frazier Norbury[a,b]

[a] Department of Psychology, Royal Holloway, University of London

[b] Division of Psychology and Language Sciences, University College London

Author Note

Correspondence concerning this article should be addressed to Courtenay Norbury, Division of Psychology and Language Sciences, University College London, Chandler House, 2 Wakefield Street, London, WC1N 1PF, United Kingdom. Electronic mail may be sent to Katie.Whiteside.2013@live.rhul.ac.uk or c.norbury@ucl.ac.uk

Abstract

Purpose: This study explored whether a monolingual-normed English language battery could identify children with English as an additional language (EAL) who have persistent English language learning difficulties that impact on functional academic attainment.

Method: Children with EAL (*n* = 43) and monolingual English-speaking children (*n* = 46) completed a comprehensive monolingual-normed English language battery in Year 1 (ages 5-6 years) and Year 3 (ages 7-8 years). Children with EAL and monolingual peers, who either met monolingual criteria for language impairment or typical development on the language battery in Year 1, were compared on language growth between Year 1 and Year 3 and on attainment in national curriculum assessments in Year 2 (ages 6-7 years).

Results: Children with EAL and monolingual peers who met monolingual criteria for language impairment in Year 1 continued to display comparably impaired overall language ability two years later in Year 3. Moreover, these groups displayed comparably low levels of academic attainment in Year 2, demonstrating comparable functional impact of their language difficulties.

Conclusions: Monolingual-normed language batteries in the majority language may have some practical value for identifying bilingual children who need support with language learning, regardless of the origin of their language difficulties.

Language impairment is a developmental disorder characterised by persistent language difficulties which have a functional impact (Reilly et al., 2014). Approximately 7.58% of monolingual children experience language impairment in the absence of an existing medical diagnosis or intellectual impairment (Norbury et al., 2016) and the prevalence is thought to be comparable in bilingual children (Kohnert, 2010). However, identifying language impairment in bilingual children is a challenge (Bedore & Peña, 2008; De Lamo White & Jin, 2011; Paradis, Schneider, & Duncan, 2013). It is recommended that language impairment is only diagnosed in bilingual children following assessment in both languages, ideally using measures which are normed on bilingual children (Bedore & Peña, 2008; RCSLT Specific Interest Group in Bilingualism, 2007). However, such measures are not available for all bilingual children, nor are they feasible to develop or administer, particularly in populations containing a high proportion of children with diverse first languages (De Lamo White & Jin, 2011). For example, in England 19.4% of state-funded primary school pupils speak English as an additional language (EAL; Department for Education, 2015), with over 300 different first languages spoken by these children (NALDIC, 2012). Due to a lack of alternative resources, practitioners generally use monolingual-normed language measures when assessing bilingual children (Caesar & Kohler, 2007; Williams & McLeod, 2012). While such methods may not accurately identify language impairment in bilingual children, it is possible that a comprehensive language battery in the majority language may identify bilingual children who experience difficulties with language learning which negatively impact academic life. To date, no research has followed the language development of bilingual children who meet criteria for language impairment on such assessment batteries or considered the functional academic attainment of those identified.

Monolingual children with language impairment vary in their language profiles (Conti-Ramsden & Crutchley, 1997). Thus, it is recommended that receptive and expressive language skills are assessed in a variety of language domains when diagnosing language impairment (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE consortium, 2016). Precise cut-offs and exclusionary criteria for language impairment are, however, under debate (Bishop, 2014; Reilly et al., 2014). Tomblin, Records, and Zhang (1996) developed the EpiSLI diagnostic system for language impairment in an epidemiological study of monolingual children. Children completed receptive and expressive measures of vocabulary, grammar, and narrative and composite scores were calculated for each modality and language domain. Children scoring -1.25 *SD* or more below the standardised mean on two out of five language composites were regarded as having language impairment. However, Norbury et al. (2016) found that a -1.5 *SD* cut on two or more composites yielded a group of children with language impairment who experienced greater functional academic impairment, relative to those identified by the -1.25 *SD* cut. With regard to exclusionary criteria, the original EpiSLI criteria required children to have normal nonverbal ability and no existing medical diagnosis. However, the requirement for a discrepancy between language and nonverbal ability is no longer endorsed by the majority of researchers and practitioners (Bishop et al., 2016), nor is it supported by epidemiological evidence (Norbury et al., 2016; Reilly et al., 2014).

Standardised monolingual-normed language measures have not been recommended for the assessment of bilingual children (Bedore & Peña, 2008). This is because typically developing bilingual children generally show poorer performance relative to monolingual peers on individual language measures, including on measures of receptive vocabulary, expressive vocabulary, receptive grammar, and narrative comprehension (Babayiğit, 2014; Bialystok, Luk, Peets, & Yang, 2010; Burgoyne, Kelly, Whiteley, & Spooner, 2009; Burgoyne, Whiteley, & Hutchinson, 2011; Hutchinson, Whiteley, Smith, & Connors, 2003; Verhoeven, Steenge, van Weerdenburg, & van Balkom, 2011). Furthermore, individual language measures can be poor at identifying bilingual children with language impairment.

In a cross-sectional study of children aged 6, 7, and 8 years, Verhoeven et al. (2011) found that measures of receptive vocabulary and narrative comprehension did not discriminate between bilingual children with language impairment and typically developing bilingual peers at ages 6 and 7 years. Moreover, measures of expressive vocabulary and receptive grammar also did not discriminate between these groups at age 6. Furthermore, both Boerma, Leseman, Timmermeister, Wijnen, and Blom (2016) and Thordardottir and Brandeker (2013) found that while typically developing bilingual children outperformed bilingual peers with language impairment on a measure of receptive vocabulary, they performed comparably to monolingual children with language impairment.

Sentence repetition tasks have been identified as a potential nonbiased measure of language in bilingual children. Sentence repetition is sensitive to language impairment in monolingual children (Conti-Ramsden, Botting, & Faragher, 2001; Riches, 2012) and is included as a measure of expressive grammar in most diagnostic batteries, including Tomblin et al.'s (1996) EpiSLI system. Monolingual and bilingual children with language impairment show comparably impaired sentence repetition accuracy (Thordardottir & Brandeker, 2013; Tsimpli, Peristeri, & Andreou, 2016). However, typically developing bilingual children also often show deficits in sentence repetition accuracy relative to typically developing monolingual peers (Chiat et al., 2013; Thordardottir & Brandeker, 2013; Tsimpli et al., 2016). Thus, sentence repetition measures may over-identify language impairment in bilingual children. Nevertheless, typically developing bilingual children differ from both monolingual and bilingual children with language impairment in their sentence repetition error patterns (Komeili & Marshall, 2013; Meir, Walters, & Armon-Lotem, 2015). These studies however used specific bilingual language groups and thus the error patterns characteristic of typical development, or language impairment, in these groups may not generalise to all bilingual children.

Narrative production tasks, which require children to generate a story or retell a previously presented story using a series of pictures, are generally considered a less-biased measure of language in bilingual children (Boerma et al., 2016; Cleave, Girolametto, Chen, & Johnson, 2010). Specifically, typically developing bilingual children do not differ from monolingual peers in narrative macro-structure, which concerns the inclusion of key story elements within the narrative (Boerma et al., 2016; Hipfner-Boucher et al., 2014; Rezzonico et al., 2015; Rodina, 2016). Moreover, monolingual and bilingual children with language impairment show comparably impaired narrative macro-structure (Boerma et al., 2016; Cleave et al., 2010; Rezzonico et al., 2015). Other studies have found that bilingual children with language impairment show poorer narrative macro-structure than typically developing bilingual peers (Paradis et al., 2013; Squires et al., 2014), though there are notable exceptions (Iluz-Cohen & Walters, 2012; Tsimpli et al., 2016). Narrative production tasks may therefore help identify language impairment in bilingual children, though it is recommended that they are used in combination with other measures (Boerma et al., 2016; Paradis et al., 2013).

While many studies have explored bilingual children's performance on individual language measures, little research has explored their performance on comprehensive diagnostic batteries. Gillam, Peña, Bedore, Bohman, and Mendez-Perez (2013) explored the diagnostic accuracy of assessing Spanish-English bilingual children in English using Tomblin et al.'s (1996) EpSLI system. All children were in first grade and had been exposed to English regularly for at least one year. Language impairment was identified using assessment in both Spanish and English. The original EpiSLI diagnostic criteria, of two or more English language composites falling -1.25 *SD* below the mean, correctly identified 95% of bilingual children with language impairment (sensitivity), though only 45% of unimpaired children were correctly identified (specificity). Adjusting the cut-offs for the individual composites yielded 86% sensitivity and 68% specificity (composite cut-offs ranged from -1.11 *SD* to -

1.83 *SD*). Moreover, acceptable sensitivity and specificity levels of 81% were yielded after combining all five composites within a predictive model. Therefore, Gillam et al. concluded that assessment in English can be used to diagnose language impairment in bilingual children who have been exposed to English regularly for at least a year. Of note, all children scored below the 30th percentile on two out of four subtests on a Spanish-English screener completed two years before the diagnostic assessment. Although Gillam et al. reported that English and Spanish skills within the sample spanned the full continua of proficiency at the time of the diagnostic assessment, the recruitment method may have biased the results.

The current study is the first to explore the persistence and functional impact of the English language difficulties experienced by children learning EAL who meet criteria for language impairment on a comprehensive monolingual-normed English language battery. Note that we do not imply that these children necessarily have an underlying language impairment; their scores on the English language battery fall in the range obtained by monolingual children with language impairment, which may reflect limited exposure to English, language impairment, or both. Children learning EAL were compared to monolingual peers, who either met criteria for language impairment or typical development on the language battery in Year 1 (ages 5-6 years). Language growth was assessed in all four groups between Year 1 and Year 3 (ages 7-8 years) and academic attainment was measured in Year 2 (ages 6-7 years). Growth is reported for a total language composite score and for the six individual language measures which make up the battery. The diagnostic battery followed the EpiSLI system, however we used a stricter cut for language impairment of two or more language composite scores falling -1.5 *SD* below the monolingual population mean. All children entered school at the same time and had received at least one year of exposure to English in school prior to the Year 1 assessment.

In contrast to Gillam et al.'s (2013) study, the children learning EAL in our population sample had diverse first languages, representing 19 languages in total. It was therefore not possible to compare the diagnostic accuracy of the language battery against language impairment diagnoses derived from dual language assessment. This is because there are no normed first language measures in most of these languages, nor would such measures be feasible to develop or administer (De Lamo White & Jin, 2011). As such, our sample represents the current educational and clinical situation in richly diverse communities and motivates our investigation to determine the use of an English language battery to assess children with EAL. To acknowledge that our English language measures are not sufficient to diagnose language impairment in children with EAL, we use the term *low language proficiency* to refer to children who met the monolingual criteria for language impairment.

Gillam et al. (2013) hypothesized that children with EAL who are typical language learners should learn English faster than those with an underlying language impairment. Following this rationale, if our language battery identified many children with EAL who do not have an underlying language impairment, we expected the EAL low language proficiency group (EAL-LL) to show greater language growth relative to the monolingual low language proficiency group (Mon-LL), and potentially greater academic attainment. Reduced persistence and functional impact of the English language difficulties experienced by the EAL-LL group, relative to the Mon-LL group, would indicate that using a monolingual-normed language battery to assess children with EAL in Year 1 may be of limited value. Alternatively, if the EAL-LL and EAL typical language proficiency (EAL-TL) groups show comparable language growth and academic attainment to their respective monolingual peers, this would indicate that the language battery may have value in identifying children with EAL who experience persistent English language learning difficulties which have functional impact. Furthermore, comparable persistence and functional impact of the English language difficulties experienced by the EAL-LL and Mon-LL groups would indicate a need for both groups to receive additional support, regardless of the origin of their difficulties.

Method

Study Design

All children were participants in the Surrey Communication and Language in Education Study (SCALES). All children who started reception year (kindergarten) in a state-maintained school in Surrey, UK, in September 2011 were eligible to take part in the first phase of SCALES (*N* = 12,398). Teachers completed an online questionnaire for 7,267 reception year children (ages 4-5 years), who attended a total of 161 state maintained schools across Surrey (59% of all eligible children; 61% of all eligible schools; see Figure 1 for recruitment details). Teachers reported that the main language spoken in the homes of 782 children (11%) was a language other than English; these children were regarded as speaking EAL.

The online questionnaire included the Children’s Communication Checklist-Short (CCC-S). The CCC-S is comprised of 13 items from the Children’s Communication Checklist-2 (CCC-2; Bishop, 2003a) which best discriminated children with language impairment from typically developing peers in a validation study (Norbury, Nash, Baird, & Bishop, 2004). Within the CCC-S, the respondent rates the frequency with which each child displays six communicative errors and seven communicative strengths. High CCC-S scores reflect greater language difficulties. Monolingual children and children with EAL scoring 1 *SD* or more above the monolingual population mean for their age group (autumn, spring, or summer born) and sex were regarded as high-risk for language impairment. All remaining children were regarded as low-risk for language impairment. Children were regarded as having *no phrase speech* (NPS) if teachers reported that the child did not produce utterances of at least two to three words. These children (62 monolingual children, 27 children with EAL) received the maximum CCC-S score of 39. A higher proportion of children with EAL

(3.45%) were reported as having NPS relative to monolingual children (0.96%; $\chi^2(1) = 35.96$, $p < .001$, Cramér's $V = .07$).

In the second phase of SCALES, subsamples of monolingual children and children with EAL were selected for in-depth assessment in Year 1 (ages 5-6 years) and Year 3 (ages 7-8 years) using stratified random sampling (see Figure 1). All children attending special schools were excluded from selection. Within the monolingual sample, high-risk children were oversampled (40.5% of screened high-risk boys selected, 37.5% of high-risk girls) relative to the low-risk children (4.3% of low-risk boys, 4.2% of low-risk girls). Within the EAL sample, a random sample of 30 high-risk (15 boys, 15 girls) and 30 low-risk (15 boys, 15 girls) children were invited to participate. Additionally, all NPS children were invited to participate (48 monolingual children, 22 children with EAL). Therefore, within the EAL sample, children with particularly low levels of English language proficiency in reception year were oversampled.

As shown in Figure 1, 529 monolingual children and 61 children with EAL, from a total of 151 state-maintained schools, completed an in-depth assessment in Year 1. Of these children, 499 monolingual children and 51 children with EAL were also assessed in Year 3. In Year 1, children were randomly assigned to one of six assessment blocks, which mapped onto the six half terms of the UK school year. In Year 3, children remained in their original assessment block, however the order of the blocks was reversed. Therefore, a child who was assessed in the first half term of Year 1 was re-assessed in the last half term of Year 3. Consequently, the lag between Year 1 and Year 3 assessments for each child varied between 14 and 34 months. This novel design maximised the longitudinal component of the study.

An opt-out consent procedure was adopted for the first phase of SCALES, in which anonymised teacher questionnaire data were submitted to the study unless parents opted out. Parents provided informed, written consent for the second phase of SCLAES, which involved in-depth, individual assessment. The study protocol was developed in collaboration with Surrey County Council and was granted ethical approval by the Ethics Committee at Royal Holloway, University of London.

(FIGURE 1 ABOUT HERE)

Participants

Figure 1 details the selection process for the current study. Children with EAL (*n* = 53) were individually matched with monolingual children (*n* = 53) after the Year 1 assessment on sex, age at assessment (within 2 months), date of birth (within 2 months), and language proficiency status in Year 1 (typical or low English language proficiency). Language proficiency status was determined using language composite scores from the English language battery (outlined below). Children with intellectual disability (i.e., those scoring 2 *SD* or more below the monolingual population mean on a nonverbal composite, outlined below), a reported medical diagnosis, and children whose language proficiency status was unclassifiable due to missing data were excluded from this matching (eight children with EAL excluded; see Figure 1). Of the matched children, nine children with EAL and five monolingual children were not assessed in Year 3 and two monolingual children and one child with EAL were excluded due to having a medical diagnosis reported in Year 3 (see Figure 1 for details).

The final sample for this study includes 43 children with EAL and 46 monolingual children who were assessed in both Year 1 and Year 3. Of this sample, 19 children with EAL (12 boys, 7 girls) and 18 monolingual children (10 boys, 8 girls) had low English language proficiency (EAL-LL; Mon-LL) in Year 1 and 24 children with EAL (11 boys, 13 girls) and 28 monolingual children (13 boys, 15 girls) had typical English language proficiency (EAL-TL; Mon-TL) in Year 1. There was no significant association between sex and language group ($\chi^2(3) = 1.75, p = .626$, Cramér's $V = .14$). All children started school at the compulsory age in the UK and thus had all received at least one year of exposure to English in school by the Year 1 assessment. The children with EAL had one of 19 different first languages. The most frequently reported first languages were Polish (7 children), Bengali (6 children), and Urdu (5 children). All other languages had three speakers or less. All children attended one of 63 state-maintained primary or infant schools in Surrey in Year 1 (children with EAL = 35 schools; monolingual children = 38 schools) and one of 61 state-maintained primary or junior schools in Surrey in Year 3 (children with EAL = 36 schools; monolingual children = 37 schools).

During assessment in Year 1, all children were aged between 5 years 3 months (63 months) and 6 years 8 months (80 months). During assessment in Year 3, all children were aged between 7 years 1 month (85 months) and 8 years 8 months (104 months). The four groups did not significantly differ in age at assessment in Year 1 or Year 3 and the lag between Year 1 and Year 3 assessments did not significantly differ between the groups (see Table 1). Income Deprivation Affecting Children Index (IDACI; McLennan et al., 2011) rank scores were retrieved using the children's home postcodes to provide a measure of neighbourhood deprivation. IDACI rank scores can range from 1 to 32,482, with lower scores assigned to areas in England with proportionally more children living in income deprived families (defined by receiving certain means tested benefits). IDACI rank scores ranged from 5,293 to 31,962 for the children with EAL and from 4,686 to 32,416 for the monolingual children, thus both groups varied widely in socioeconomic backgrounds. The four groups did not significantly differ in IDACI rank scores (see Table 1).

(TABLE 1 ABOUT HERE)

Measures and Procedures

Each child completed an individual two hour assessment session with a trained researcher when they were in Year 1 and Year 3. Assessment sessions took place in a quiet area in each child's school and were broken up with breaks. All tasks were administered in English. The measures relevant to this study included assessments of nonverbal ability and language. This study also incorporates data from national curriculum assessments, provided by Surrey County Council, which were completed when the children were in Year 2.

Nonverbal ability. In Year 1, children completed the Block Design and Matrix Reasoning subtests of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III; Wechsler, 2003b). In Year 3, children completed the Block Design and Matrix Reasoning subtests of the Wechsler Intelligence Scale for Children (WISC-IV; Wechsler, 2003a). At each time point, raw scores on these tasks were converted into age standardised nonverbal ability composite *z* scores, based on norms derived from the monolingual population sample.

Language. In Year 1 and Year 3, children completed the Receptive and Expressive One-Word Picture Vocabulary Tests, Fourth Edition (ROWPVT-4, EOWPVT-4; Martin & Brownell, 2011a, 2011b). Children also completed a narrative production (expressive) and comprehension (receptive) task based on the Narrative subtest of the Assessment of Comprehension and Expression (ACE 6-11; Adams, Cooke, Crutchley, Hesketh, & Reeves, 2001). Moreover, children completed the Test for Reception of Grammar - Short (TROG-S), which is a short form of the Test for Reception of Grammar-2 (TROG-2; Bishop, 2003b), and the School Age Sentence Imitation Task - English 32 (SASIT-E32; Marinis, Chiat, Armon-Lotem, Piper, & Roy, 2011) to assess receptive and expressive grammar, respectively.

Raw scores on the six language measures from the Year 1 and Year 3 assessments were converted into age standardised *z*-scores based on norms derived from the monolingual population sample. Following Tomblin, Records, and Zhang (1996), five composite scores were calculated: vocabulary, grammar, narrative, expressive language, and receptive language. Low language proficiency in Year 1 was defined as two or more language composite scores falling -1.5 *SD* or more below the monolingual population mean, in the absence of existing medical diagnoses or intellectual disability (defined as a nonverbal ability composite score of -2 *SD* or more below the monolingual population mean). This criteria has been used to identify language impairment in monolingual English-speaking children (Norbury et al., 2016). For both time points, z-scores on all six language measures were averaged to produce a total language composite *z*-score.

Receptive vocabulary (ROWPVT-4; Martin & Brownell, 2011b). The examiner read individual words and children were asked to select a picture, from an array of four, which depicted each word. Scores ranged from 0-190, with higher scores indicating greater receptive vocabulary.

Expressive vocabulary (EOWPVT-4; Martin & Brownell, 2011a). Children were presented with a series of individual pictures and were asked to name the object, action, or concept which was depicted in each picture. Scores ranged from 0-190, with higher scores indicating greater expressive vocabulary.

Narrative recall (ACE 6-11; Adams et al., 2001). Children listened to a story about a monkey in a forest, which was played over headphones and accompanied by eight pictures. After listening to the story, children were shown the pictures again and were asked to tell the story in their own words. Each child's narrative was audio-recorded and 1 point was awarded for each of 35 key elements of the story which were correctly recalled (maximum = 35).

Narrative comprehension. Following the narrative production task, children were asked 12 comprehension questions about the story (six literal and six inference questions). Children received 0 points for an incorrect response, 1 point for a partially correct response, and 2 points for a correct response (maximum = 24).

Receptive grammar (TROG-S). This is a short form of the TROG-2 (Bishop, 2003b). Children heard up to 40 sentences and were asked to select a picture, from an array of four, which depicted each sentence. The task was discontinued if a child answered incorrectly on six consecutive items. One point was allocated for each correct response (maximum = 40).

Sentence repetition (SASIT-E32; Marinis et al., 2011). Children listened to 32 pre-recorded sentences over headphones and were asked to repeat each sentence out load. All repetitions were audio-recorded and 1 point was allocated for every sentence that was repeated correctly (word for word; maximum = 32).

Year 2 assessments. Children attending state-maintained schools in England complete national curriculum assessments, known as Key Stage 1 assessments, in the last term of Year 2 (ages 6-7 years; Department for Education, 2014). Teachers determine each child's level of attainment in the following five subjects: mathematics, science, reading, writing, and speaking and listening. Since the expected level of attainment is level 2 (Department for Education, 2014), children were regarded as performing on target in each subject if they achieved level 2 or above.

Data Analysis

All analyses were conducted using Stata IC 14 (StataCorp, 2015). Two-way independent measures ANOVAs, with EAL status and language proficiency status as independent variables, tested whether nonverbal ability composite z-scores significantly differed between the language groups in Year 1 and Year 3. Pearson's correlations between raw scores achieved in Year 1 and Year 3 on each language measure, as well as correlations between Year 1 and Year 3 total language composite z-scores, are provided separately for children with EAL and monolingual children. Chi-square tests indicated whether children from the EAL-TL and EAL-LL groups differed from monolingual peers in their likelihood to perform on target in all five subjects in Year 2 assessments (versus performing below target in one or more subject) and perform on target in each individual subject.

A series of linear mixed effects models, with robust standard errors, were run to explore the relationship between language group membership and growth, or stability, between Year 1 and Year 3 in total language composite z-scores and raw scores on each language measure. While models using raw scores show the change in actual scores over time, models using z-scores show whether the groups differ in their performance relative to the monolingual population sample over time. Linear mixed effects modelling was appropriate as this analysis allows unequal testing intervals. Moreover, linear models allow an estimate of language growth despite only two testing observations. Child ID was entered into each model as a random effect to account for individual variation at initial assessment (the intercept). EAL status (EAL, monolingual), language proficiency (typical language proficiency, low language proficiency), and age in months (continuous) were entered into each model as fixed effects. The following interaction terms were also entered into each model: Language Proficiency x EAL, Language Proficiency x Age, EAL x Age, and Language Proficiency x EAL x Age.

Within each model, coefficients reveal the relationship between each variable and the outcome when all other variables are at 0. Age was centred at the mean age for all participants during testing in Year 1 (71.34 months). Thus, for age, 0 reflects the mean age in Year 1. For language proficiency, 0 reflects typical language proficiency and for EAL status, 0 reflects being monolingual. Coefficients for language proficiency therefore reveal how the Mon-LL group compares to the Mon-TL group in Year 1 and coefficients for EAL status reveal how the EAL-TL group compares to the Mon-TL group in Year 1. The Language Proficiency x EAL interaction reveals whether the relationship between learning EAL and the outcome in Year 1 differs for children with low language proficiency relative to children with typical language proficiency (i.e., the extent the difference between EAL-LL and Mon-LL groups is comparable to the difference between EAL-TL and Mon-TL groups). Coefficients for age reveal the relationship between age in months and the outcome variable (i.e. growth in the outcome, or the slope) for the Mon-TL group. The Language Proficiency x Age interaction reveals whether growth is different for the Mon-LL group relative to the Mon-TL group. The EAL x Age interaction reveals whether growth is different for the EAL-TL group relative to the Mon-TL group. Finally, the Language Proficiency x EAL x Age interaction reveals whether the difference in growth between EAL-LL and Mon-LL groups is comparable to the difference in growth between EAL-TL and Mon-TL groups.

For linear mixed models which demonstrated a significant interaction involving EAL status and age, a second linear mixed model considered performance in Year 3, with age centred at the mean age at assessment in Year 3 (95.34 months). The models were built in the same way as the original models in all other respects. For these models, coefficients are only reported for language proficiency, EAL status, and the Language Proficiency x EAL interaction. Coefficients for age and the interactions involving age are identical to the original models. For models which had no significant interactions involving EAL status and age, no further analyses were undertaken. Such a result indicates that the disparity between children with EAL and monolingual peers that was evident in Year 1 remained over time.

Missing Data

Two children (both EAL-LL) did not complete the SASIT-E32 in Year 1 and one child (EAL-TL) did not complete the SASIT-E32 in Year 3. As these children did not complete the full language battery, they were excluded from the models predicting total language composite z-scores, as well as the models predicting sentence repetition. Two children (1 EAL-LL, 1 EAL-TL) did not complete the WISC-IV Matrix Reasoning subtest in Year 3 and were excluded from the nonverbal ability analysis in Year 3.

Results

Means and *SDs* for Year 1 and Year 3 nonverbal ability z-scores, total language composite z-scores, and raw scores on each language measure are displayed in Table 2 for each group. A two-way ANOVA demonstrated a significant main effect of language proficiency on nonverbal ability composite *z*-scores in Year 1, $F(1, 85) = 4.97$, $p = .028$, $\eta_p^2 = .06$, and Year 3, $F(1, 83) = 30.57$, $p < .001$, $\eta_p^2 = .27$, whereby children with typical language proficiency had higher scores relative to children with low language proficiency. However, there was no significant main effect of EAL status in Year 1, $F(1, 85) = 0.20$, $p = .655$, $\eta_p^2 < .01$, or Year 3, $F(1, 83) = 1.51$, $p = .223$, $\eta_p^2 = .02$, nor was there a significant EAL x Language Proficiency interaction in Year 1, $F(1, 85) = 0.07$, $p = .786$, $\eta_p^2 < .01$, or Year 3, $F(1, 83) = 0.42$, $p = .520$, $\eta_p^2 = .01$. Thus, within language proficiency groups, children with EAL did not differ from monolingual peers in nonverbal ability in Year 1 or Year 3.

For both children with EAL and monolingual children, Year 1 and Year 3 total language composite z-scores, and raw scores on each language measure, were significantly positively correlated (see Table 3).

(TABLE 2 AND 3 ABOUT HERE)

Growth in Total Language Composite Z-scores

As displayed in Table 4 and Figure 2, when age was mean centred in Year 1, language proficiency status significantly predicted total language composite z-scores. Furthermore, EAL status did not significantly predict total language composite z-scores, and there was no significant Language Proficiency x EAL interaction. Thus, as expected, both Mon-LL and EAL-LL groups obtained lower total language composite z-scores relative to Mon-TL and EAL-TL groups in Year 1. Moreover, EAL-TL and EAL-LL groups achieved comparable total language composite z-scores in Year 1 to their respective monolingual peer groups. Age in months did not significantly predict total language composite z-scores, which indicates that total language composite z-scores remained constant for the Mon-TL group as age increased. There was a significant Language Proficiency x Age interaction, a marginally significant EAL x Age interaction, but no significant Language Proficiency x EAL x Age interaction. Thus, Mon-LL and EAL-LL groups made greater growth in total language composite z-scores relative to Mon-TL and EAL-TL groups, respectively (see Figure 2). Moreover, EAL-TL and EAL-LL groups made slightly greater growth in total language composite z-scores relative to their respective monolingual peer groups.

A separate mixed linear model, with age mean centred in Year 3, demonstrated that language proficiency status in Year 1 continued to significantly predict total language composite z-scores in Year 3, $\beta = -1.34$, $p < .001$, 95% CI [-1.76, -0.93]. Additionally, EAL status did not significantly predict total language composite z-scores, $\beta = 0.18$, $p = .462$, 95% CI [-0.31, 0.68], and there was no significant Language Proficiency x EAL interaction, $\beta = -0.27$, $p = .415$, 95% CI [-0.91, 0.38]. Thus, despite greater growth, both Mon-LL and EAL-LL groups continued to display lower total language composite z-scores relative to Mon-TL and EAL-TL groups in Year 3. Furthermore, EAL-TL and EAL-LL groups continued to demonstrate comparable total language composite z-scores to their respective monolingual peer groups in Year 3. Both EAL-LL and Mon-LL groups performed on average at least 1.5 *SD* below the monolingual population mean in Year 1 and Year 3 (see Figure 2).

(FIGURE 2, TABLE 4, TABLE 5, & FIGURE 3 ABOUT HERE)

Growth in Raw Scores on Each Language Measure

Tables 4 and 5 display linear mixed models predicting raw scores on each language measure when age was mean centred in Year 1. Language proficiency status significantly predicted scores on each language measure. Thus, the Mon-LL group obtained significantly poorer scores on all measures relative to the Mon-TL group in Year 1 (see Figure 3). EAL status did not significantly predict receptive vocabulary, narrative recall, and receptive grammar and there were no significant Language Proficiency x EAL interactions for any of these measures. Therefore, EAL-TL and EAL-LL groups displayed comparable receptive vocabulary, narrative recall, and receptive grammar in Year 1 to their respective monolingual peer groups. EAL status also did not significantly predict expressive vocabulary and narrative comprehension, however there was a significant Language Proficiency x EAL interaction for both of these measures. Thus, while EAL-TL and Mon-TL groups displayed comparable expressive vocabulary and narrative comprehension in Year 1, the EAL-LL group performed more poorly on these measures relative to the Mon-LL group (see Figure 3). Finally, EAL status significantly predicted sentence repetition, with no significant Language Proficiency x EAL interaction. Thus, both EAL-TL and EAL-LL groups displayed poorer sentence repetition relative to their respective monolingual peer groups in Year 1.

Age significantly predicted scores on each measure, which indicates that the Mon-TL group displayed growth in all measures over time. For expressive vocabulary, narrative recall, and receptive grammar, there were no significant interactions involving age. Thus, all groups displayed comparable growth in expressive vocabulary, narrative recall, and receptive grammar and the rank order of the performance by the groups in Year 1, on these measures, remained in Year 3 (see Figure 3). Specifically, EAL-TL and EAL-LL groups continued to display comparable narrative recall and receptive grammar to their respective monolingual peer groups. EAL-TL and Mon-TL groups also continued to display comparable expressive vocabulary and the EAL-LL group continued to display lower expressive vocabulary than the Mon-LL group.

For narrative comprehension, there was a significant Language Proficiency x Age interaction, but no significant EAL x Age or Language Proficiency x EAL x Age interaction. Therefore, both Mon-LL and EAL-LL groups demonstrated greater growth in narrative comprehension relative to Mon-TL and EAL-TL groups (see Figure 3). Furthermore, both EAL-TL and EAL-LL groups demonstrated comparable growth relative to their respective monolingual peer groups. Consequently, as evident in Year 1, EAL-TL and Mon-TL groups continued to display comparable narrative comprehension in Year 3, while the EAL-LL group continued to display poorer narrative comprehension relative to the Mon-LL group.

The Language Proficiency x Age interaction term was marginally significant for receptive vocabulary and significant for sentence repetition. Additionally, for both receptive vocabulary and sentence repetition, there was a significant EAL x Age interaction, but no significant Language Proficiency x EAL x Age interaction. Therefore, both Mon-LL and EAL-LL groups made greater growth in receptive vocabulary and sentence repetition relative to Mon-TL and EAL-TL groups, respectively (see Figure 3). Furthermore, both EAL-TL and EAL-LL groups made greater growth in receptive vocabulary and sentence repetition relative to their respective monolingual peer groups.

To investigate the EAL x Age interaction for receptive vocabulary, age was mean centred in Year 3. Language proficiency status significantly predicted receptive vocabulary, $\beta = -11.91$, $p < .001$, 95% CI [-18.74, -5.08]. Additionally, EAL status did not significantly predict receptive vocabulary, $\beta = 2.03$, $p = .509$, 95% CI [-3.99, 8.05], and there was no significant Language Proficiency x EAL interaction, $\beta = -4.95$, $p = .265$, 95% CI [-13.67, 3.76]. Therefore, Mon-LL and EAL-LL groups continued to display poorer receptive vocabulary relative to Mon-TL and EAL-TL peer groups in Year 3 (see Figure 3). Furthermore, despite greater growth, EAL-TL and EAL-LL groups continued to demonstrate comparable receptive vocabulary to their respective monolingual peer groups in Year 3.

Similarly, language proficiency status significantly predicted sentence repetition in Year 3, $\beta = -10.26$, $p < .001$, 95% CI [-13.49, -7.03]. However, EAL status did not significantly predict sentence repetition, $\beta = -1.73$, $p = .293$, 95% CI [-4.96, 1.50], nor was there a significant Language Proficiency x EAL interaction, $\beta = -0.34$, $p = .892$, 95% CI [-5.20, 4.53]. Therefore, Mon-LL and EAL-LL groups continued to display poorer sentence repetition relative to Mon-TL and EAL-TL peer groups in Year 3 (see Figure 3). However, in contrast to Year 1, EAL-TL and EAL-LL groups showed comparable sentence repetition to their respective monolingual peer groups in Year 3 (see Figure 3).

Academic Attainment

As shown in Table 6, a greater proportion of children within the Mon-TL and EAL-TL groups performed on target in all five subjects in Year 2 assessments, as well as on target in each individual subject, relative to EAL-LL and Mon-LL groups. Moreover, as shown in Table 6, for both language proficiency groups, there was no significant association between EAL status and overall attainment, or attainment in specific subjects. Therefore, EAL-TL and EAL-LL groups showed comparable attainment in Year 2 assessments relative to their respective monolingual peer groups.

(TABLE 6 ABOUT HERE)

Discussion

The current UK-based longitudinal study explored whether a monolingual-normed English language battery, administered in Year 1 (ages 5-6 years), could identify children with EAL who had persistent English language learning difficulties that impact on functional academic attainment. Comparisons were made between the language development and academic attainment of children with EAL and monolingual peers, who either had typical language development or met criteria for language impairment on the English language battery in Year 1. Despite showing moderately greater growth between Year 1 and Year 3, EAL-TL and EAL-LL groups did not significantly differ in overall language ability in Year 1 or Year 3 from their respective monolingual peer groups. Both EAL-LL and Mon-LL groups showed persistent language difficulties, performing on average at least 1.5 *SD* below the monolingual population mean on the full language battery at each time point. Furthermore, the EAL-LL group did not outperform the Mon-LL group on any individual language measure and indeed they showed particular difficulty relative to Mon-LL peers in expressive vocabulary, narrative comprehension (both Years 1 and 3), and in sentence repetition (Year 1 only). With regard to functional impact, EAL-TL and EAL-LL groups showed comparable attainment in national curriculum assessments in Year 2 relative to their respective monolingual peer groups. Therefore, monolingual criteria for language impairment on an English language battery identified children with EAL and monolingual peers who showed persistent English language learning difficulties, which were accompanied by comparable academic underachievement.

It is typically recommended that bilingual children with suspected language impairment are assessed in both of their languages, ideally using bilingual-normed measures (Bedore & Peña, 2008; RCSLT Specific Interest Group in Bilingualism, 2007). However, in populations containing a high proportion of children with diverse first languages, such as the UK, standardised first language measures are simply not available for all bilingual children, nor are they feasible to develop or administer (De Lamo White & Jin, 2011). Therefore, practitioners generally use monolingual-normed language measures when assessing bilingual children (Caesar & Kohler, 2007; Williams & McLeod, 2012). While many studies have indicated that bilingual children are often disadvantaged relative to monolingual peers on individual language measures (Babayiğit, 2014; Bialystok et al., 2010; Burgoyne et al., 2009, 2011; Hutchinson et al., 2003; Verhoeven et al., 2011), there is limited evidence to support the accuracy of diagnostic decisions based on comprehensive language diagnostic batteries. Findings from the current study suggest that a comprehensive, monolingual-normed English language battery may have some practical value for identifying children with EAL who require targeted support to develop English language proficiency.

This work extends early investigation by Gillam et al. (2013), who explored the diagnostic accuracy of assessment in English, using Tomblin et al.'s (1996) EpSLI model, to identify language impairment in Spanish-English bilingual children, who had been exposed to English daily for at least a year. Gillam et al. found that combing all five English language composites in a predictive model yielded more acceptable diagnostic accuracy (81%

sensitivity, 81% specificity) than the original EpiSLI criteria, of two or more composites falling -1.25 *SD* below the mean, which yielded many false-positives (95% sensitivity, 45% specificity). In the current study a stricter cut-off of -1.5 *SD* below the monolingual population mean on two or more composites was used and we took a novel, longitudinal approach to assessing the long-term utility of this cut-off. Since the EAL-LL group had marginally greater growth in overall language ability relative to the Mon-LL group, a proportion of children in the EAL-LL group may be false-positives. This is because children with EAL who are typical language learners should learn English faster than those with language impairment (Gillam et al., 2013). However, despite greater growth, EAL-LL and Mon-LL groups did not differ significantly in overall language ability in Year 1 or Year 3 and both groups performed on average at least 1.5 *SD* below the monolingual population mean at each time point. Thus, while we cannot be sure of the origins of the language difficulties experienced by the children in the EAL-LL group, these children experienced persistent English language difficulties over the early school years at a level comparable to their monolingual peers. Furthermore, children within the EAL-LL and Mon-LL groups achieved comparable attainment in Year 2 national curriculum assessments. These findings suggest that the English language battery has some practical value for identifying children with EAL who may benefit from targeted support, regardless of the origin of their language difficulties.

In the current study, the EAL-TL group had comparable receptive and expressive vocabulary, narrative comprehension, and receptive grammar to the Mon-TL group in Year 1 and Year 3. These findings appear to contradict research which found that typically developing bilingual children tend to achieve lower scores than monolingual peers on measures of these abilities (Babayiğit, 2014; Bialystok et al., 2010; Burgoyne et al., 2009, 2011; Hutchinson et al., 2003; Verhoeven et al., 2011). While findings from this study could be interpreted as suggesting that these measures are not biased against typically developing children with EAL, the findings may reflect that the children were compared on the same tasks which were used to form the language groups. The EAL-TL group may have thus included children with particularly high levels of English language proficiency. Nevertheless, the findings highlight that many children with EAL perform comparably to monolingual peers on standardised language measures.

The EAL-LL group in the current study had comparable receptive vocabulary and receptive grammar to the Mon-LL group in Year 1 and Year 3. In contrast, the EAL-LL group performed more poorly relative to the Mon-LL group on measures of narrative comprehension and expressive vocabulary at both time points, suggesting that these are areas of particular difficulty for children learning EAL who start school with limited English proficiency. Both EAL-TL and EAL-LL groups had poorer sentence repetition accuracy relative to their respective monolingual peer groups in Year 1. However, both EAL-TL and EAL-LL groups displayed greater growth in sentence repetition accuracy relative to the monolingual groups and by Year 3, they did not significantly differ from their respective monolingual peer groups. These findings indicate that measures of sentence repetition accuracy may be biased against children with EAL, particularly in the early school years. Thus, assessment at school entry using a measure of sentence repetition accuracy may identify many false-positives, whose poor scores reflect lack of facility with English grammar, rather than a fundamental deficit in language learning. These results are somewhat consistent with studies reporting that typically developing bilingual children show impaired sentence repetition accuracy relative to typically developing monolingual peers (Thordardottir & Brandeker, 2013; Tsimpli et al., 2016). The greater growth in sentence repetition accuracy among children with EAL, relative to monolingual peers, may reflect increased exposure to English as the children progress through school, as sentence repetition accuracy is positively associated with language exposure (Thordardottir & Brandeker, 2013).

In the current study, there was no effect of EAL status within either language proficiency group on narrative recall. There was, however, an effect of language proficiency, whereby both EAL-LL and Mon-LL groups included fewer key story elements in their narratives relative to EAL-TL and Mon-TL groups. This is consistent with studies that have reported no effects of bilingualism on narrative macro-structure among children with typical development (Boerma et al., 2016; Hipfner-Boucher et al., 2014; Rezzonico et al., 2015; Rodina, 2016) or language impairment (Boerma et al., 2016; Cleave et al., 2010; Rezzonico et al., 2015). Findings are also consistent with reports that language impairment is associated with impaired narrative macro-structure in monolingual and bilingual children (Boerma et al., 2016; Cleave et al., 2010; Paradis et al., 2013; Rezzonico et al., 2015; Squires et al., 2014). Thus, findings from this study support the assertion that narrative recall tasks are a nonbiased measure of language ability in bilingual children (Boerma et al., 2016; Cleave et al., 2010).

The study has a number of strengths relative to previous investigations. The children were recruited from a population sample, therefore the sample is not biased towards particular language or cultural communities and is representative of children learning EAL in the UK. In contrast, previous studies on language impairment in bilingual children have typically recruited children from specific language communities (e.g. Spanish or French) and have selected children from clinical caseloads or specialist schools, which introduces bias. Another major strength of this study is the longitudinal design. This allowed the persistence of language difficulties experienced by children with EAL, and monolingual peers, to be compared over the early school years. Moreover, the unique design of this study, which resulted in the lag between the Year 1 and Year 3 assessments varying between 14 and 34 months, maximized the longitudinal element of the study. Finally, the incorporation of data from national curriculum assessments provided an ecologically valid measure of functional impact at school. While an increasing body of literature has explored how to distinguish language impairment from limited language experience in bilingual children, none of this research, to the authors knowledge, has considered functional impact.

However, our conclusions are tempered by limitations of the study. Assessing the children at more time points would have further strengthened this study by enabling us to determine whether the EAL-LL and Mon-LL groups continue to show comparable language ability and academic attainment, or whether the EAL-LL group do eventually catch up to the EAL-TL group. Moreover, as there were only two time points, the current analyses were limited to linear growth models. Additional time points would allow the investigation of potential non-linear growth trajectories. Such designs also allow for analysis of latent growth profiles, which may provide a complementary assessment by identifying children who demonstrate sustained improvement, versus those with more stable patterns of language deficit. The long term implications of using such English (or majority) language assessment could therefore be evaluated. These are important avenues for future research.

This study is also limited by the lack of data on exposure to English. Nevertheless, since all children in the current study started school at the mandatory age, we know that all children had been exposed to English for at least one year in school by the Year 1 assessment and for at least three years by the Year 3 assessment. We also know that 98% of children in the local area take advantage of government-funded nursery provision (15 hours per week from age three; Surrey County Council, Early Years Team, personal communication, 2015), suggesting that the majority of children had received regular exposure to English from age three. It should also be noted that since the children with EAL in the current study were recruited at school entry, they have only ever experienced an English school environment. As a result, the findings from the current study concerning comparisons with monolingual peers, and the predictive ability of the English language battery, may not be applicable to children with EAL with more variable backgrounds, such as children who join an English school during a later stage in their education.

Another point of consideration is that 10 children within the current study, eight of whom were within the EAL-LL group, were reported to have no phrase speech (NPS) in reception year, whereas only two children from the monolingual sample were reported to have NPS in reception year. This is consistent with the higher proportion of children with EAL, relative to monolingual children, who were reported to have NPS in the population survey phase. Our study is unable to determine whether NPS status in reception year in the EAL sample reflects more limited exposure to English prior to school entry, or is indicative of an underlying language disorder. More detailed information about home language environment and family history of language learning impairment is needed to distinguish these possibilities. Oversampling children with NPS in the EAL sample may have yielded an EAL-LL group with more persistent language learning challenges. Nevertheless, the EAL and monolingual groups in the current study were matched according to English language performance in Year 1.

The lack of assessment of first language proficiency is also a limitation. This would have allowed an investigation of the proportion of the EAL-LL group who also experienced difficulties in their first language, giving an indication of the specificity of the diagnostic criteria used in the study. Nevertheless, we argue that this is not a practical goal. In this study, 19 different first languages were represented and over 300 different first languages are represented by school children in the UK (NALDIC, 2012). It is unlikely that robust diagnostic instruments will be available at any point in the near future for all of these languages. Thus, the investigation of English language tools that aid identification of children who need support with language learning and academic achievement, remains an important endeavour.

While the current study used a language battery comprised of six language measures from multiple publishers, practitioners may consider whether findings will hold if alternative language measures are used or, indeed, if a language battery such as the Clinical Evaluation of Language Fundamentals (CELF-4; Semel, Wiig, & Secord, 2003) is used. To the extent the language assessment taps receptive and expressive skills in multiple language domains, has demonstrated long-term stability in monolingual cohorts, and uses comparable cut-offs to the current study, one would expect the findings to hold. Nevertheless, it is important for future research to explore the long-term utility of other language batteries for the assessment of bilingual children.

In conclusion, the current UK-based longitudinal study found that criteria for language impairment on a monolingual-normed English language battery, administered in Year 1, identified children with EAL and monolingual children who showed persistent English language difficulties over the early school years, which were accompanied by a comparable academic impact. We cannot be certain that the children with EAL who were identified using the battery have an underlying language impairment. However, the findings indicate that these children may require additional targeted support, regardless of their origins of their language difficulties. Therefore, monolingual-normed language batteries in the majority language may have some practical value for assessing bilingual children in populations where first language measures are not available.

Acknowledgements

This research was supported by a grant awarded by the Wellcome Trust (WT094836AIA) to C.F.N. and a Crossland Scholarship awarded by Royal Holloway, University of London, to K.E.W. We would like to thank the schools, parents, and children who took part, as well as Surrey County Council for providing data from Year 2 assessments. The views expressed in this paper are those of the authors and not necessarily those of the Wellcome Trust or Surrey County Council. The authors have no conflict of interests.

References

Adams, C., Cooke, R., Crutchley, A., Hesketh, A., & Reeves, D. (2001). *Assessment of Comprehension and Expression 6-11 (ACE 6-11)*. London, United Kingdom: GL assessment.

Babayiğit, S. (2014). The role of oral language skills in reading and listening comprehension of text: a comparison of monolingual (L1) and bilingual (L2) speakers of English language. *Journal of Research in Reading*, *37*(S1), S22–S47. https://doi.org/10.1111/j.1467-9817.2012.01538.x

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 1–29. https://doi.org/10.2167/beb392.0

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bishop, D. V. M. (2003a). *Children's Communication Checklist – 2*. London, United Kingdom: Pearson.

Bishop, D. V. M. (2003b). *Test for Reception of Grammar-Version 2 (TROG-2)*. London, United Kingdom: Pearson Assessment.

Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language & Communication Disorders*, *49*(4), 381–415. https://doi.org/10.1111/1460-6984.12101

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PeerJ Preprints*. https://doi.org/10.7287/peerj.preprints.1986v1

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language & Communication Disorders*, n/a-n/a. https://doi.org/10.1111/1460-6984.12234

Burgoyne, K., Kelly, J. M., Whiteley, H. E., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *British Journal of Educational Psychology*, *79*(4), 735–747. https://doi.org/10.1348/000709909X422530

Burgoyne, K., Whiteley, H., & Hutchinson, J. M. (2011). The development of comprehension and reading-related skills in children learning English as an additional language and their monolingual, English-speaking peers. *British Journal of Educational Psychology*, *81*(2), 344–354. https://doi.org/10.1348/000709910X504122

Caesar, L. G., & Kohler, P. D. (2007). The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools*, *38*(3), 190–200. https://doi.org/10.1044/0161-1461(2007/020)

Chiat, S., Armon-Lotem, S., Marinis, T., Polišenská, K., Roy, P., & Seeff-Gabriel, B. (2013). The potential of sentence imitation tasks for assessment of language abilities in sequential bilingual children. In V. Mueller-Gathercole (Ed.), *Issues in the assessment of bilinguals* (pp. 56–89). United Kingdom: Mulitlingual Matters. Retrieved from http://openaccess.city.ac.uk/4797/1/

Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. *Journal of Communication Disorders*, *43*(6), 511–522. https://doi.org/10.1016/j.jcomdis.2010.05.005

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Conti-Ramsden, G., & Crutchley, A. (1997). The extent to which psychometric tests differentiate subgroups of children with SLI. *Journal of Speech, Language & Hearing Research*, *40*(4), 765. https://doi.org/10.1044/jslhr.4004.765

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders*, *46*(6), 613–627. https://doi.org/10.1111/j.1460-6984.2011.00049.x

Department for Education. (2014). *Phonics screening check and key stage 1 assessments: England 2014* (No. SFR 34/2014). Retrieved from https://www.gov.uk/government/statistics/phonics-screening-check-and-key-stage-1-assessments-england-2014

Department for Education. (2015). *Schools, pupils and their characteristics: January 2015* (No. SFR 16/2015). Retrieved from https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2015

Gillam, R. B., Peña, E. D., Bedore, L. M., Bohman, T. M., & Mendez-Perez, A. (2013). Identification of specific language impairment in bilingual children: I. Assessment in English. *Journal of Speech, Language & Hearing Research*, *56*(6), 1813–1823. https://doi.org/10.1044/1092-4388(2013/12-0056)

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2014). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism*, 1367006914534330. https://doi.org/10.1177/1367006914534330

Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading*, *26*(1), 19–32. https://doi.org/10.1111/1467-9817.261003

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, *15*(Special Issue 01), 58–74. https://doi.org/10.1017/S1366728911000538

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, *43*(6), 456–473. https://doi.org/10.1016/j.jcomdis.2010.02.002

Komeili, M., & Marshall, C. R. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clinical Linguistics & Phonetics*, *27*(2), 152–162. https://doi.org/10.3109/02699206.2012.751625

Marinis, T., Chiat, S., Armon-Lotem, S., Piper, J., & Roy, P. (2011). *School Age Sentence Imitation Task - English 32 (SASIT-E32)*. Retrieved from http://www.city.ac.uk/health/research/centre-for-language-communication-sciences-research/veps-very-early-processing-skills/veps-assessments

Martin, N. A., & Brownell, R. (2011a). *Expressive One-Word Picture Vocabulary Test, Fourth Edition*. Novato, CA: Academic Therapy Publications.

Martin, N. A., & Brownell, R. (2011b). *Receptive One-Word Picture Vocabulary Test, Fourth Edition*. Novato, CA: Academic Therapy Publications.

McLennan, D., Barnes, H., Noble, M., Davies, J., Garratt, E., & Dibben, C. (2011). *The English Indices of Deprivation 2010*. London, United Kingdom: Department for Communities and Local Government. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6320/1870718.pdf

Meir, N., Walters, J., & Armon-Lotem, S. (2015). Disentangling SLI and bilingualism using sentence repetition tasks: the impact of L1 and L2 properties. *International Journal of Bilingualism*, 1367006915609240. https://doi.org/10.1177/1367006915609240

NALDIC. (2012). Languages in schools. Retrieved from http://www.naldic.org.uk/research-and-information/eal-statistics/lang

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., … Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry. Advance Online Publication*. https://doi.org/10.1111/jcpp.12573

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. M. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: A validation of the Children's Communication Checklist — 2. *International Journal of Language & Communication Disorders*, *39*(3), 345–364. https://doi.org/10.1080/13682820410001654883

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, *56*(3), 971–981. https://doi.org/10.1044/1092-4388(2012/12-0050)

RCSLT Specific Interest Group in Bilingualism. (2007). *Good practice for speech and language therapists working with clients from linguistic minority communities*. RCSLT London, UK. Retrieved from http://www.rcslt.org/members/publications/publications2/linguistic_minorities

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., … Wake, M. (2014). Specific language impairment: a convenient label for whom? *International Journal of Language & Communication Disorders*, *49*(4), 416–451.

Rezzonico, S., Chen, X., Cleave, P. L., Greenberg, J., Hipfner-Boucher, K., Johnson, C. J., … Girolametto, L. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI. *International Journal of Language & Communication Disorders*, *50*(6), 830–841. https://doi.org/10.1111/1460-6984.12179

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language & Communication Disorders*, *47*(5), 499–510. https://doi.org/10.1111/j.1460-6984.2012.00158.x

Rodina, Y. (2016). Narrative abilities of preschool bilingual Norwegian-Russian children. *International Journal of Bilingualism*, 1367006916643528. https://doi.org/10.1177/1367006916643528

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals - Fourth Edition (CELF-4)*. San Antonio, TX: The Psychological Corporation.

Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language & Communication Disorders*, *49*(1), 60–74. https://doi.org/10.1111/1460-6984.12044

StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of*

Communication Disorders, *46*(1), 1–16. https://doi.org/10.1016/j.jcomdis.2012.08.002

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1284–1294. https://doi.org/10.1044/jshr.3906.1284

Tsimpli, I. M., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, *37*(Special Issue 01), 195–216. https://doi.org/10.1017/S0142716415000478

Verhoeven, L., Steenge, J., van Weerdenburg, M., & van Balkom, H. (2011). Assessment of second language proficiency in bilingual children with specific language impairment: A clinical perspective. *Research in Developmental Disabilities*, *32*(5), 1798–1807. https://doi.org/10.1016/j.ridd.2011.03.010

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children - Fourth Edition (WISC - IV)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003b). *Wechsler Preschool and Primary Scales of Intelligence - Third UK Edition*. London, United Kingdom: Pearson Assessment.

Williams, C. J., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology*, *14*(3), 292–305. https://doi.org/10.3109/17549507.2011.636071

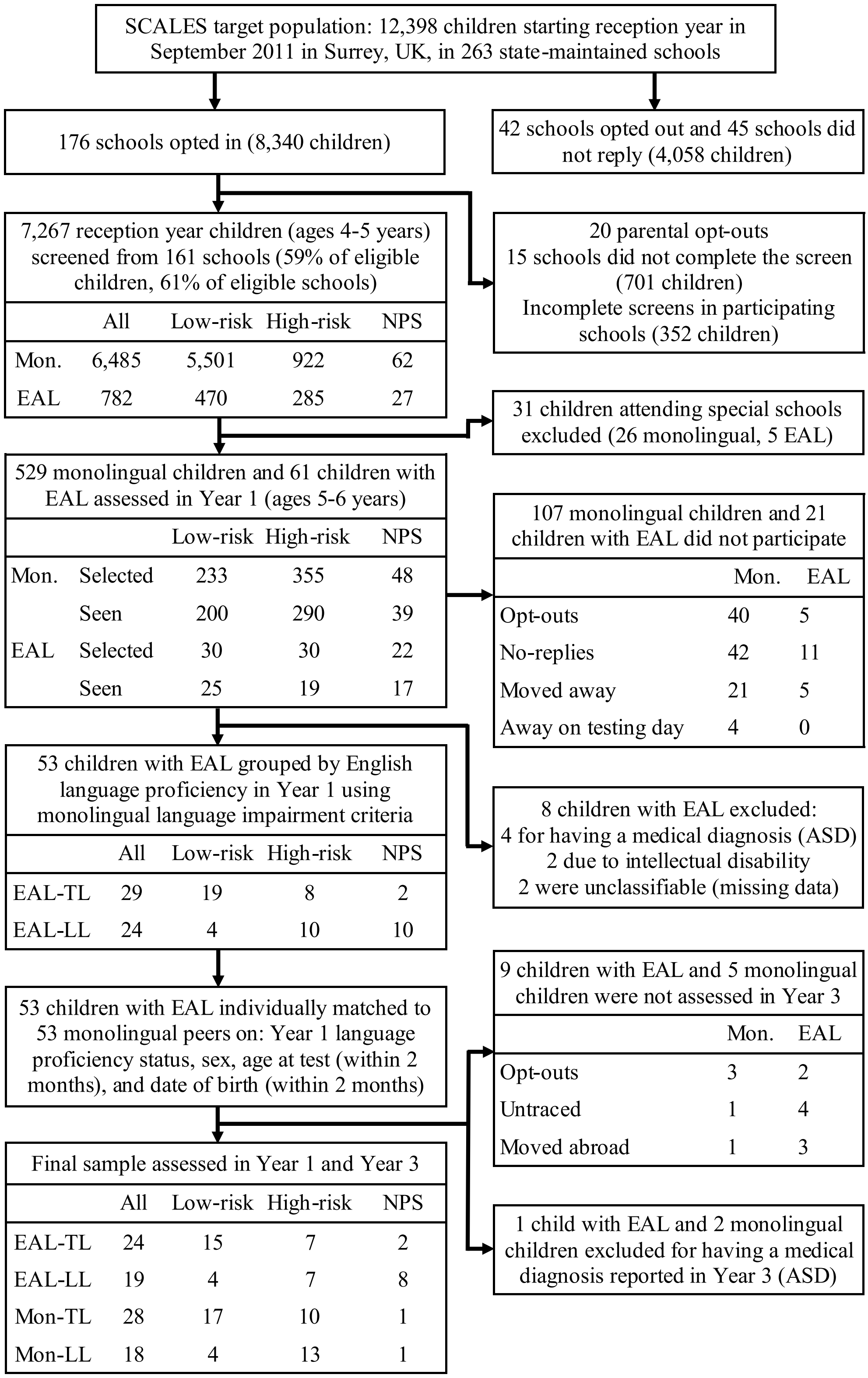

Figure 1. Recruitment flow chart for the population survey phase, Year 1 school assessment, and the selection and retention of participants in the current study. NPS = no phrase speech; ASD = autism spectrum disorder; Mon. = monolingual; EAL = English as an additional language; TL = typical language; LL = low language.

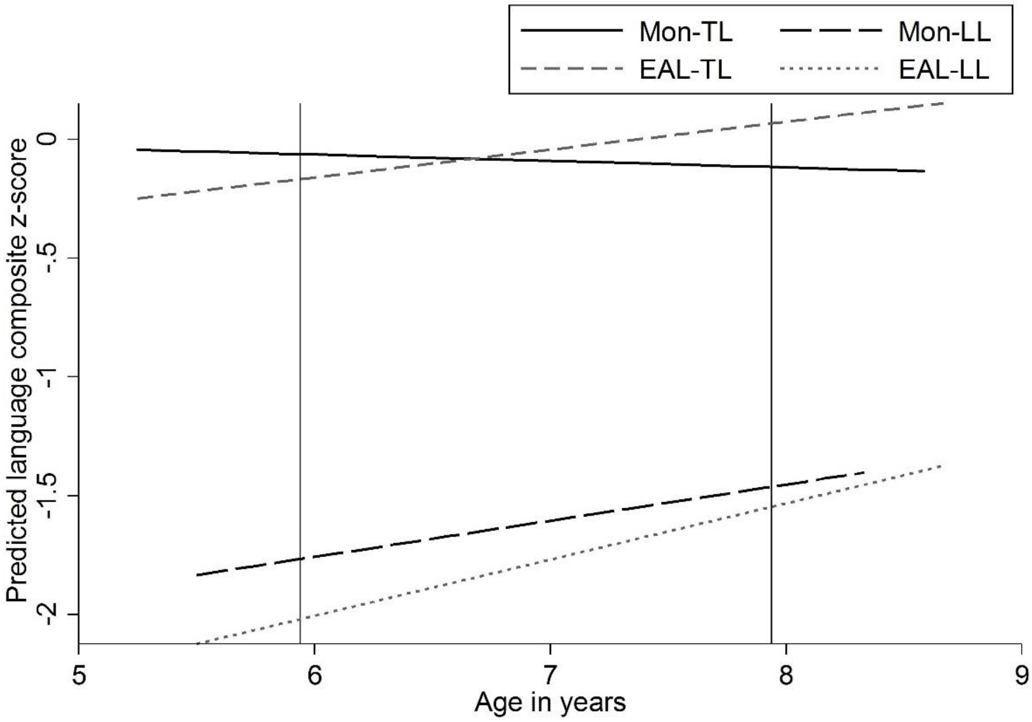

Figure 2. Predicted total language composite z-scores for each language group. The reference lines indicate the mean ages during testing in Year 1 and Year 3.

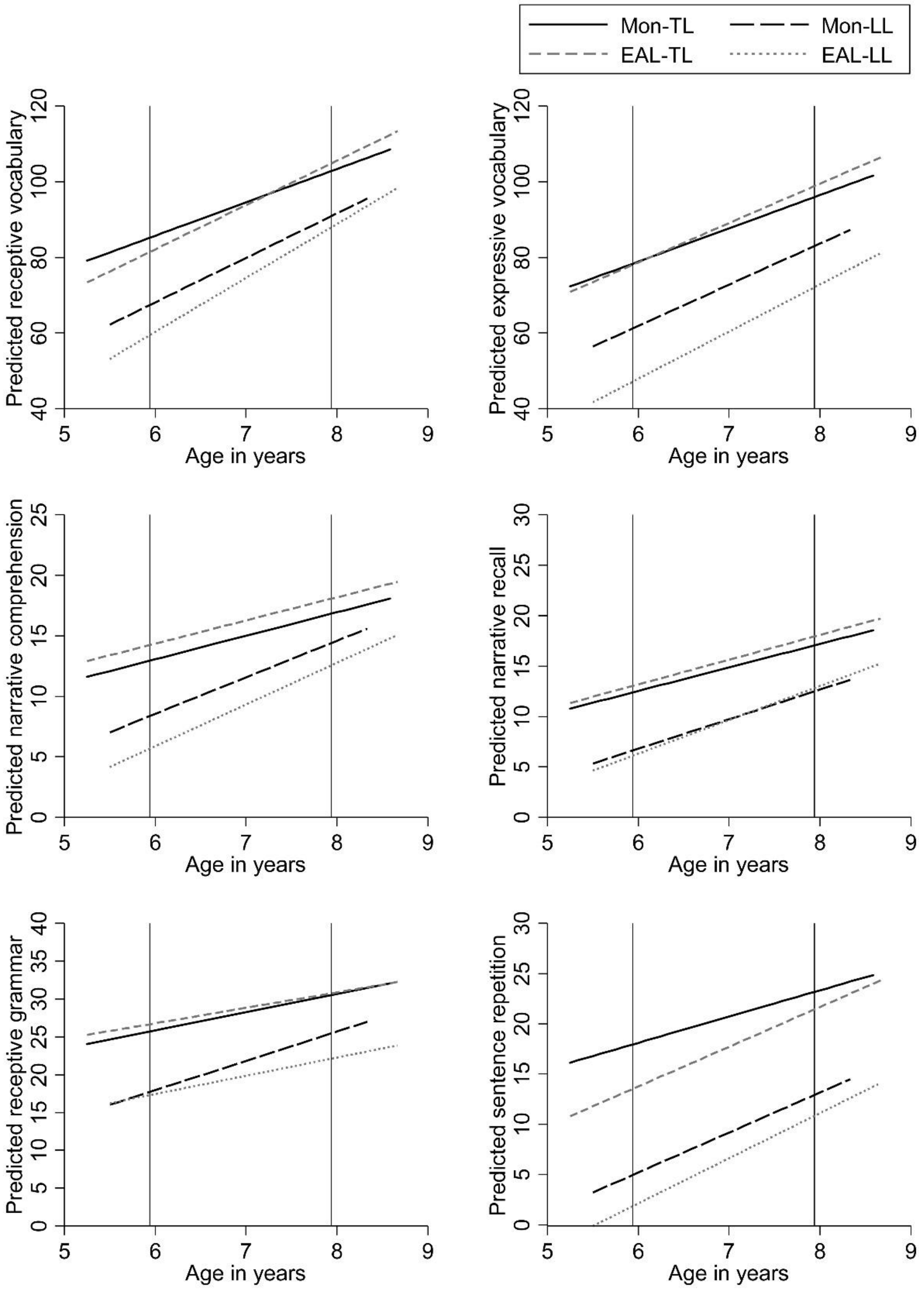

Figure 3. Predicted raw scores on measures of receptive and expressive vocabulary, narrative, and grammar for each language group. The reference lines indicate the mean ages during testing in Year 1 and Year 3.

Table 1

Mean (SD) IDACI Rank Scores, Age at Both Time Points, and Lag Between Assessments for Each Language Group

Measure	Mon-TL (n = 28)	Mon-LL (n = 18)	EAL-TL (n = 24)	EAL-LL (n = 19)	*F (df)*	*p*	η_p^2
IDACI rank	21445.29 (6403.21)	18943.17 (8211.84)	18938.54 (8394.40)	16061.84 (9048.24)	1.75 (3, 85)	.163	.06
Age in Year 1 in months	70.96 (4.02)	72.17 (4.15)	70.83 (4.30)	71.74 (4.25)	0.49 (3, 85)	.693	.02
Age in Year 3 in months	95.54 (4.24)	94.83 (4.09)	95.67 (4.45)	95.11 (5.27)	0.15 (3, 85)	.928	.01
Months between assessments	24.57 (5.09)	22.67 (5.20)	24.83 (5.36)	23.37 (6.24)	0.73 (3, 85)	.534	.03

Table 2

Mean (SD) Nonverbal Ability and Total Language Composite Z-scores, and Raw Scores on Each Language Measure, for Each Language Group in Year 1 and Year 3

	Mon-TL		Mon-LL		EAL-TL		EAL-LL	
Measure	Year 1	Year 3	Year 1	Year 3	Year 1	Year 3	Year 1	Year 3
Nonverbal composite	-0.12 (1.19)	0.17 (0.90)	-0.66 (0.82)	-0.90 (0.86)	-0.08 (0.97)	0.27 (0.77)	-0.50 (0.89)	-0.57 (0.54)
Language composite	-0.09 (0.93)	-0.09 (0.83)	-1.77 (0.45)	-1.46 (0.63)	-0.20 (0.77)	0.10 (1.00)	-1.99 (0.42)	-1.58 (0.66)
Receptive vocabulary	84.50 (12.99)	103.54 (11.55)	68.56 (14.50)	90.22 (11.72)	80.42 (9.52)	105.83 (9.92)	59.89 (8.29)	87.89 (12.36)
Expressive vocabulary	77.86 (13.49)	96.43 (11.78)	62.00 (9.91)	82.67 (10.13)	77.50 (14.05)	99.38 (10.68)	48.05 (16.02)	71.53 (16.36)
Narrative comprehension	12.82 (4.07)	16.93 (2.83)	8.33 (3.20)	14.56 (3.05)	13.92 (4.21)	18.38 (2.99)	5.79 (4.01)	12.53 (5.46)
Narrative recall	12.18 (4.12)	17.25 (3.38)	6.72 (2.74)	12.50 (4.12)	12.79 (4.11)	18.17 (3.24)	6.16 (3.62)	12.84 (4.80)
Receptive grammar	25.64 (6.14)	30.61 (5.40)	18.00 (4.95)	25.33 (3.74)	26.58 (5.06)	30.83 (4.42)	17.37 (5.92)	22.11 (7.04)
Sentence repetition	17.71 (6.96)	23.39 (5.72)	5.22 (3.44)	12.78 (5.11)	13.48 (8.00)	21.57 (5.51)	2.18 (3.30)	10.47 (6.65)

Table 3

Longitudinal Correlations Between Year 1 and Year 3 Total Language Composite Z-scores and Raw Scores on Each Language Measure for Children with EAL and Monolingual Peers

	Monolingual		EAL	
Measure	*r*	*p*	*r*	*p*
Language composite	.84	< .001	.84	< .001
Receptive vocabulary	.74	< .001	.60	< .001
Expressive vocabulary	.75	< .001	.77	< .001
Narrative comprehension	.57	< .001	.70	< .001
Narrative recall	.44	.002	.51	.001
Receptive grammar	.52	< .001	.69	< .001
Sentence repetition	.85	< .001	.75	< .001

Table 4

Linear Mixed Models Predicting Growth in Total Language Composite Z-scores, Receptive Vocabulary, Expressive Vocabulary, and Narrative Comprehension

	Language composite z-score		Receptive vocabulary		Expressive vocabulary		Narrative comprehension	
Fixed factor	β [95% CI]	*p*	β [95% CI]	*p*	β [95% CI]	*P*	β [95% CI]	*p*
LP	-1.70 [-2.08, -1.32]	< .001	-17.79 [-24.63, -10.96]	< .001	-17.13 [-23.15, -11.10]	< .001	-4.57 [-6.50, -2.63]	< .001
EAL	-0.10 [-0.56, 0.36]	.657	-3.77 [-9.26, 1.72]	.178	-0.30 [-7.24, 6.63]	.932	1.29 [-0.90. 3.48]	.247
LP x EAL	-0.15 [-0.69, 0.39]	.589	-4.17 [-12.93, 4.59]	.351	-13.73 [-24.46, -2.99]	.012	-3.99 [-7.05, -0.93]	.011
Age	< -0.01 [-0.01, 0.01]	.605	0.73 [0.57, 0.90]	< .001	0.73 [0.58, 0.89]	< .001	0.16 [0.11, 0.21]	< .001
LP x Age	0.01 [< -0.01, 0.03]	.016	0.25 [-0.01, 0.50]	.056	0.18 [-0.03, 0.38]	.091	0.09 [0.01, 0.17]	.028
EAL x Age	0.01 [< -0.01, 0.02]	.051	0.24 [0.02, 0.46]	.029	0.13 [-0.08, 0.35]	.225	< -0.01 [-0.08, 0.08]	.965
LP x EAL x Age	-0.01 [-0.03, 0.02]	.647	-0.03 [-0.41, 0.34]	.865	-0.01 [-0.37, 0.36]	.978	0.04 [-0.10, 0.17]	.586

Note. LP = language proficiency. Age is measured in months and was centred at the mean age at assessment in Year 1 (71.34 months).

Table 5

Linear Mixed Models Predicting Growth in Narrative Recall, Receptive Grammar, and Sentence Repetition

	Narrative recall		Receptive grammar		Sentence repetition	
Fixed factor	β [95% CI]	*p*	β [95% CI]	*p*	β [95% CI]	*p*
LP	-5.77 [-7.69. -3.85]	< .001	-7.99 [-10.93, -5.06]	< .001	-12.99 [-15.76, -10.21]	< .001
EAL	0.65 [-1.39, 2.68]	.534	0.94 [-1.94, 3.82]	.521	-4.42 [-8.39, -0.44]	.030
LP x EAL	-1.16 [-4.00, 1.68]	.425	-1.39 [-5.69, 2.91]	.526	1.32 [-3.11, 5.74]	.560
Age	0.19 [0.14, 0.25]	< .001	0.20 [0.10, 0.30]	< .001	0.22 [0.15, 0.29]	< .001
LP x Age	0.05 [-0.06, 0.16]	.351	0.12 [-0.02, 0.26]	.089	0.11 [0.01, 0.21]	.027
EAL x Age	0.01 [-0.07, 0.09]	.818	-0.03 [-0.13, 0.07]	.569	0.11 [0.01, 0.21]	.029
LP x EAL x Age	0.02 [-0.13, 0.18]	.754	-0.09 [-0.29, 0.11]	.381	-0.07 [-0.23, 0.09]	.407

Note. LP = language proficiency. Age is measured in months and was centred at the mean age at assessment in Year 1 (71.34 months).

Table 6

The Number (Percentage) of Children Within Each Language Group Who Performed on Target in Each Subject in Year 2 Assessments. Chi-Square Tests Explored Attainment by EAL Status Separately For Groups With Typical Language Proficiency And Groups With Low Language Proficiency

Subject	Mon-TL (*n* = 28)	EAL-TL (*n* = 24)	χ^2 (*df* = 1)	*P*	*V*	Mon-LL (*n* = 18)	EAL-LL (*n* = 19)	χ^2 (*df* = 1)	*P*	*V*
All subjects	23 (82%)	22 (92%)	1.01	.316	.14	9 (50%)	9 (47%)	0.03	.873	.03
Science	26 (93%)	24 (100%)	1.78	.182	.19	12 (67%)	12 (63%)	0.05	.823	.04
Maths	25 (89%)	24 (100%)	2.73	.099	.23	12 (67%)	15 (79%)	0.71	.401	.14
Writing	25 (89%)	23 (96%)	0.78	.377	.12	11 (61%)	10 (53%)	0.27	.603	.09
Reading	26 (93%)	24 (100%)	1.78	.182	.19	11 (61%)	14 (74%)	0.67	.414	.13
Speaking and listening	24 (86%)	23 (96%)	1.52	.217	.17	11 (61%)	12 (63%)	0.02	.898	.02