CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Collocational knowledge in children: a comparison of English-speaking monolingual children, and children acquiring English as an Additional Language

Nick RICHES[1] [iD], Carolyn LETTS[1], Hadeel AWAD[1], Rachel RAMSEY[2], and Ewa DĄBROWSKA[2]

[1]Newcastle University, UK and [2]Northumbria University, UK
Address for correspondence: Nick Riches, University of Newcastle – Education Communication and Language Studies, Newcastle upon Tyne, NE1 7RU nick.riches@newcastle.ac.uk

Abstract

Collocations, **e.g.,** apples and pears, hard worker, constitute an important avenue of linguistic enquiry straddling both grammar and the lexicon. They are sensitive to language experience, with adult L2 learners and children learning English as an Additional Language (EAL) exhibiting poor collocational knowledge. The current study piloted a novel collocational assessment with children (mean age 6;3, 40 monolingual, 32 EAL). It investigated (1) the feasibility of a collocational assessment at this age, (2) whether collocational knowledge is associated with other language domains (receptive grammar and vocabulary), and (3) whether collocational knowledge is more affected than other domains. The assessment demonstrated good psychometric properties and was highly correlated with performance in other domains, indicating shared psycholinguistic mechanisms. Unlike adult counterparts, the EAL children performed equally poorly across domains. Given the role played by collocations in vocabulary development and reading, a focus on this domain may be beneficial for EAL children.

Keywords: English as an additional language; collocations; syntax

Introduction

A key aspect of language is its generativity, which enables us to combine words in novel ways to express novel meanings, e.g., *odourless smelly concepts slumber viscously* (adapted from Chomsky, 1957), or *she coughed the napkin off the* table (adapted from Goldberg, 1995). Though the term "generative" is frequently associated with Chomskyan Generative Linguistics, all theories of linguistic knowledge must account for this property. However, in addition to creating novel utterances, speakers also

CrossMark

recycle formulaic units. These range from collocations, or 'words which go together' such as *make amends,* or *hazard a guess*, to phrasal idioms such as *kick the bucket* (= to die), or even entire sentences, e.g., *what's a nice girl like you doing in a place like this?* Such formulaic units vary substantially in terms of their size and behaviour. While some are "frozen", in that they disallow lexical additions or substitutions, e.g., *kick the pail* loses the idiomatic meaning of *kick the bucket*, others contain open slots for arguments, e.g., *she read* [*her neighbour / her partner*] *the riot act*. Some are idiomatic, e.g., *kick the bucket*, while others are compositional, e.g., *hazard a guess.* Despite this variability, all formulaic utterances share the common property that they are likely to be stored and produced as wholes (Wray, 2000, p. 465; as cited by Gibbs, 2010). A key means of determining whole storage is collocational probability, whether the likelihood of words co-occurring is greater than would be predicted from their individual frequencies.

There is considerable debate regarding the role of formulaic units within the language system. The generative system is vital for expressing novel meanings, to convey one's communicative intentions, and this can be viewed as a fundamental property of language. This aspect of language has been emphasised by Chomskyan linguistic theory. However, formulaic units are also communicatively important. For example, they can be used to regulate social interaction (e.g., topic-switching formulae such as *I was wondering… about that X…*) or express agreement or disagreement (e.g., *Tell me about it, No way!*). Chunks that are informal or used by a narrow speech community (e.g., *nuff said, nae bother*) can be used to express social identify or foster group cohesion. The use of chunks facilitates processing. Because they may be retrieved as wholes, they buy the speaker time to generate more complex sequences. Given the communicative and processing importance of formulaic language, Wray and Perkins (2000) have proposed a "division of labour" account, arguing that individuals with a high degree of linguistic competence are able to manipulate and coordinate both the generative and formulaic systems. According to Wray (1998, p. 64) "without the rule-based system, language would be limited in repertoire, clichéd, and, whilst suitable for certain types of interaction, lacking imagination and novelty. In contrast, with only a rule-based system, language would sound pedantic, unidiomatic and pedestrian."

While Wray's account depicts two competing systems, it is arguably not clear where the dividing line should be placed. For example, the formulaic expression *she read* [*her neighbour / her partner*] *the riot act* is lexicalised, with specific words occurring in specific slots; but it is also syntactic, with an open slot for the insertion of arguments, and the potential to modify tense and aspect on the verb. This duality has led some researchers in the fields of usage-based or cognitive linguistics to propose that distinctions between domains are graded, rather than absolute, and that domain-general learning and memory mechanisms support linguistic representations from across different domains, e.g., syntax, the lexicon and phonology (Bybee, 2010; Langacker, 2009). However, others have sought to keep domains fundamentally separate. For example, Culicover, Jackendoff and Audring (2017) note that multi-word constructions are almost exclusively grammatical, and this well-formedness results from both the specification of internal structure in the lexicon, and the licensing of these constructions by the grammatical system. Thus, a formal division between the lexical and grammatical systems is maintained.

A second controversy relates to the role which formulaic language plays in language development. Many researchers have argued that a repository of unanalysed

multi-word, or multi-morpheme units, provides the raw materials for acquiring abstract morphosyntactic rules or principles. Marchman and Bates (1994) have argued that we need a "critical mass" of around fifty unanalysed verb forms in order to acquire inflections, e.g., the past tense marker. Dąbrowska and Lieven (2005) have likewise argued that the acquisition of more complex structures, e.g., English interrogatives, depends on a similar process of analysing formulaic units. Usage-based theories argue that language learners, both children and adults, may acquire language via gradual abstraction from rote-learned formulae (Ellis, 2003; Tomasello, 2003). Indeed, child language gives the appearance of being highly formulaic. For example, Pine and Lieven (1997) have argued that children tend to use particular determiners (definite and indefinite articles) with particular nouns, indicating an absence of an abstract determiner category. Nonetheless, the data are subject to interpretation. For example, Yang, Crain, Berwick, Chomsky and Bolhuis, (2017) have argued that apparent evidence for formulaic units in corpora may arise from the statistical properties of these corpora rather than underlying psycholinguistic processes.

In summary, there are two main controversies regarding the function of formulaic language within the broader linguistic system: firstly, its relationship to other linguistic domains, and secondly, its role in language development. In order to investigate these issues, an important first step is to devise a test of individual knowledge of formulaic multi-word units. Surprisingly, research in this area is sparse. While there have been a number of studies assessing children's ability to interpret and explain idioms (e.g., Cain, Towse & Knight, 2009; Nippold & Martin, 1989), few studies have assessed individual differences in collocational knowledge. To our knowledge, there are three studies by Dąbrowska and colleagues (Dąbrowska, 2014, 2019; Llompart & Dąbrowska, 2020) which have investigated this construct in adults, and only one, by Smith and Murphy (2015), which studied children of primary-school age.

So far, the data indicate a close relationship between collocational knowledge and other linguistic domains. Dąbrowska (2014) presented a word with five possible collocates within a multiple choice context, e.g., *deliver/hold/perform/present/utter a speech*, and asked individuals which combination sounded "the most natural or familiar". Collocations were either Verb + Noun (as above) or Adjective + Noun (e.g., *urgent matters*, *small details*) The authors found that collocational knowledge was significantly associated with both receptive lexical and grammatical measures (Dąbrowska, 2014, 2019; Llompart & Dąbrowska, 2020). In the most recent study, which investigated 60 English-speaking adults, these associations were strong: 0.80 for the association with grammar, and 0.82 for the association with vocabulary (Llompart & Dąbrowska, 2020). These relationships were even stronger than the relationship between grammar and vocabulary (0.70). Likewise, Smith and Murphy tested knowledge of Verb + Noun collocations, e.g., *pay attention*, *change direction*, in primary-school aged children (children aged 7–10 with a mean age of 8.43 years). They also employed a written format, but for their task both collocates were chosen from a choice of three. They also found a strong association between collocational abilities, and vocabulary scores, e.g., an association of r = 0.59 with the British Picture Vocabulary Scales (Dunn, Dunn & Styles, 2009).

Dąbrowska *et al.* claim that their findings support "the essential premise of usage-based models of language acquisition that all linguistic knowledge is represented in the same format, as pairings between form and meaning" (p. 11). Given these representational similarities the same learning and memory mechanisms are assumed to be exploited across domains. However, they also find that

collocational knowledge is influenced by cognitive traits such as phonological short-term memory, assessed using the digit span task, and measures of explicit language learning, e.g., a linguistic rule induction task, and a task which involves learning object-label associations. These tasks arguably tap into domain-specific processes. For example, they rely to a greater or lesser extent on phonological short-term memory, to store a phonological representation of a numeric sequence, or a novel word form. Thus, although it is conceivable that the high correlations reflect the operation of domain-general systems, there may also be a wide range of cognitive factors, some domain-specific, which impact on performance across linguistic systems.

Smith and Murphy's discussion of correlations is focused on the strong relationship between the collocations task, and the ACE (Assessment of Comprehension and Expression) test of phrasal verbs (e.g., *put up*, *show off*) (Adams, Cooke, Crutchley, Hesketh & Reeves, 2001). Such verbs themselves consist of word combinations with strong co-occurrence frequencies, and therefore it is hardly surprising that they are strongly associated with collocational knowledge. However, there are also psycholinguistic accounts which explain why collocational abilities might impact on the acquisition of single words. Often the collocates of a word will provide cues to meaning. For example, the verb *trudge* tends to co-occur with the preposition *through* implying effortful movement through a three-dimensional substance (Dabrowska, 2009). An individual who more closely attends to collocates will be more sensitive to these semantic cues.

While the above discussion focuses on cognitive factors, Dąbrowska also highlights a key environmental factor: exposure to the target language. Native English speaker scores on her "words that go together" task correlate strongly with measures of print exposure (Dąbrowska, 2014, 2019) and in nonnative speakers they are strongly associated with use of English (Dąbrowska, 2019). These data are corroborated by a study by Verhagen (2019), who tested knowledge of collocations appearing in job advertisements amongst adults with varying degrees of exposure to this genre: professional recruiters, job-seekers, and non-jobseekers. Performance on a sentence completion task, which investigated both adherence to the target sentence and voice onset time, was strongly associated with the relative experience of the groups. Collocations are especially frequency sensitive because they do not allow for robust form-function mappings. For example, though we tend to say *urgent matters*, other combinations, e.g., *important / pressing matters*, express the same concept and are also well-formed. There is evidence that children are able to rapidly acquire idiosyncratic form-function mappings via the process of fast-mapping, a process which may operate for both words, and syntactic constructions (Carey & Bartlett, 1978; Casenhiser & Goldberg, 2005). Because non-idiomatic collocations cannot be fast-mapped their acquisition is more strongly influenced by frequency.

If collocational knowledge is more sensitive to exposure than other domains, we would expect individuals who lack experience with the target language to obtain low scores on collocations tests. This is indicated by existing data. When comparing native and nonnative speakers (Dąbrowska, 2019) effect sizes were much larger for the "words which go together" test ($d = 1.61$) than the vocabulary and receptive grammar tasks (0.77 and 0.70). Such a finding is consistent with the odd collocational patterns produced by adult language learners, e.g., *make attention* (Nesselhauf, 2003). These errors persist even when sentence structure and use of individual words is highly accurate. In fact, it has been argued that non-native-like use of collocations "ultimately marks out the advanced L2 learner as non-native"

rather than grammatical or word-level lexical errors (Wray & Perkins, 2000, p. 2, referring to research by Pawley & Syder, 1983). Dąbrowska's (2019) study is the first to demonstrate that such difficulties extend beyond production.

A similar uneven profile was observed by Llompart and Dąbrowska (2020). They compared high and low educational attainment adults, and found that, similar to Dąbrowska (2019), the collocations task most strongly differentiated the groups. This resulted in a significant interaction between linguistic construct (vocabulary, grammar, collocations) and task (p < 0.01). One possible cause for this profile is the limited print exposure of low educational attainment adults. This is especially important for the "words which go together" task which sourced its items from a collocations dictionary which was based on written sources.

Collocations constitute an important but much neglected avenue of linguistic inquiry. They are theoretically important as they problematise a neat separation of linguistic domains, thereby challenging modular accounts. They also have implications for understanding how language systems develop. For example, as outlined above, there are grounds for assuming that collocational knowledge may impact on syntactic and lexical development. Other causal pathways may exist. For example, Smith and Murphy (2015) propose that collocational knowledge impacts on reading fluency – for example, by decreasing the time spent decoding non-literal phrases. However, we currently lack good tests of collocational knowledge for young children. Though Smith and Murphy's assessment exhibits strong convergent validity (correlations with other language measures), due to its written format, it is targeted at older primary-age children. We therefore explored the feasibility of testing collocational knowledge in young children. If Smith and Murphy are correct that collocational knowledge impacts longitudinally on other language domains, e.g., reading, it is important to devise a test which is sensitive to this construct at earlier ages – for example, at the onset of formal education – in order for teachers and other educational practitioners to identify and help children with a weakness in this area.

Secondly, we aimed to determine whether the correlational analyses conducted by Dąbrowska in adults and Smith and Murphy on older primary school children would be replicated in a younger group. We assumed that the main cognitive and environmental factors, e.g., working memory and variations in language exposure, which drive variation across linguistic domains in adults and older children would also be present in younger children. Finally, like Smith and Murphy (2015) and Dąbrowska (2019) we incorporated a group of children learning English as an Additional Language (EAL). We hypothesised that, due to their limited English exposure they would exhibit an uneven profile characterised by especially poor performance on the collocations task.

To recap, our aims are the following: (1) to investigate the feasibility of an assessment of collocational knowledge in young children; (2) investigate the relationship between collocational knowledge and other language subdomains; (3) explore the language profiles (including collocational knowledge) of monolingual versus EAL children.

Methods
Stimuli

Stimuli were generated through a multi-stage process. Initially, items were generated by brain-storming, with frequencies checked against corpora. Five types of collocations were used; binomials (two words of the same class linked by a coordinating

conjunction, e.g., *fish and chips, big and small*), Verb + Noun collocations (*make the bed*), Verb + Adjective collocations (*go crazy*), Adjective + Noun collocations (*hard worker*) and similes following the structure *as* + Adjective + *as* + Noun, or Verb + *like* + Noun (e.g., *as black as ink, it fits like a glove*). To facilitate selection of binomials, the thousand most frequent in the British National Corpus were identified (British National Corpus, Version 2, 2002). The corpus was accessed via an online portal hosted by Brigham Young University (Davies, 2004). We then excluded collocations where the second word contained more syllables than the first word, so that children were unable to use a word-length heuristic (Benor & Levy, 2006). All collocations were easy to depict, given the visual presentation format. In total, during this initial stage, 58 binomials were selected, along with 10 Adjective + Noun collocations, 2 Verb + Adjective collocations, 9 Verb + Noun collocations, and 28 similes (107 items in total).

The items were subsequently piloted via Reddit Sample Size, a forum dedicated to recruiting participants for online research. This was considered essential as research has problematised the status of expert intuitions (Schutze, 1996). 96 respondents (mean age 27.2 years, 20 British English speakers, 66 American English speakers and 20 speakers of another variety) chose the item which sounded "better" from a choice of two. A third option was available if both items sounded equally good. For the binomials, the distractor consisted of the same words in reverse order, e.g., *fish and chips / chips and fish*. For the other items, distractors were designed to sound plausible, and be semantically similar to the targets, e.g., *messy eater / *writer, go crazy / *sad, as black as ink / * a crow*. A score reflecting collocational strength was created by averaging responses across participant (1 = identifies target, 0 = identifies distractor, 0.5 = not sure) and converting to a percentage. Collocations were selected if this score went above 50%. The average rating of the chosen items (percentage of responses identifying the target) was 88.1%, with a standard deviation of 11.8%. The lowest rating was 50.4% (*shoes and socks*). Though a number of items were selected with relatively low ratings, e.g., four items had scores between 50% and 60%, weakly collocational items may improve the sensitivity of the assessment by minimising ceiling effects. Verhagen's (2019) data suggests that such items may be better for assessing individual differences. The final assessment contained 47 binomials, 9 V + N collocations, 2 V + A collocations, 7 A + N collocations, and 24 similes (89 items in total).

The stimuli were presented in a booklet. Each page contained a target and distractor picture, presented left to right in a randomised order. For the binomials, e.g., *hat and gloves*, each picture consisted of two images: hat and gloves. Their position reflected their order of occurrence in the binomial. For example, in the target, the hat appeared on the left, and the gloves appeared on the right (see example in Appendix). This order was reversed for the distractor picture. As the experimenter said the items, they pointed at the images from left to right, e.g., *hat* (points to image on the left) *and gloves* (points to image on the right). This ensured that the order of Nouns (and Adjectives / Verbs) within the collocation was visually represented. For the other collocations, the first part of the collocation was printed at the top of the page, e.g., "do" + A PICTURE OF DAMAGE / A PICTURE OF SOME MESS, "as black as" + A PICTURE OF INK / A PICTURE OF A CROW. The experimenter said the word at the top of the page, then produced the collocations and the distractor while pointing at the pictures from left to right, e.g., *do* (points to picture on the left) *some damage*, *do* (points to the picture on the right) *a mess*.

The testing materials have been deposited in the Open Science Framework (https://osf.io/) under the first author's name, and are also provided as supplementary materials (Supplementary Materials).

Participants

35 EAL children and 40 age-matched monolingual children were recruited from 6 primary schools in the North East of England chosen for their large populations of EAL children. The participants were from Year One, which contains children who turn six in the current school year. The mean age for both groups was 6;3 (see below). The schools were asked to send invitation letters to parents/caregivers of children in mainstream classes who were acquiring English as an L1 or additional language. Parents/caregivers of EAL children were sent a paper questionnaire to determine the child's EAL status. This asked questions about the child's language background containing the following fields (exact wording); 'Name of child', 'Age (in years / months)', 'Sex (Male / Female)', 'Age when the child first came to the UK', 'Child's first language', 'Age when child first started hearing or speaking English on a regular basis', 'Does your child have a medical difficulty which might affect their language?' In addition, parents/caregivers were asked about their first language and how frequently they spoke English with the child ('always', 'usually', 'half the time', 'rarely', 'never').

Children were classified as EAL if they did not have a native English-speaking parent, and/or their age of regular exposure to English was 3;0 or above. No parent reported a medical difficulty affecting language. The children's L1s were predominantly non-European (Arabic, Bengali, Bangla, Bini, Berber, 'Chinese', Dari, Farsi, Kurdish, Mandarin, Urdu), with only 3 speaking a European L1 (German, Czech, Portuguese). A further two children were included who were identified as EAL by the school, but whose family did not return the questionnaire. Teachers' judgements of EAL status were highly accurate, with only 1 out of 36 children identified as EAL not meeting the study criteria according to the questionnaire.

For the EAL group, the mean age of regular exposure to English (current age minus age of first regular exposure to English) was 49.2 months. The standard deviation (20.35) was large, reflecting substantial variation. 13 children were exposed to English outside the home within the first year, while 9 children were not exposed to English until their fifth year (4;0 and above). This relative heterogeneity should be borne in mind when interpreting the data.

Procedure

The collocations task was framed as a detective game. Children were told they would hear some phrases, some of which were 'good', while others were 'not so good'. It was their job as a language detective to decide which sounded 'better'. To signal their choice they could either repeat the 'better' phrase, or point to the associated picture. In each trial, the experimenter read aloud the two expressions (target and distractor), and simultaneously pointed to the corresponding picture. For binomials, they pointed at each image from left to right, as described above.

The child was familiarised with the task using three training trials, in which the distractors contained very obvious word order errors, e.g., *kick ball / ball kick*. In the first two trials, the question "which one sounds better?" was used to prompt a

response. In the final training trial, if children appeared to understand the task, the question cue was not used. After the training phase, the child was asked if they wished to continue. If, during the testing phase, children deviated from the task, they were reminded of the procedure. A short break of around one minute was administered half way through.

To investigate the children's language profiles, tests of receptive grammar and vocabulary were administered. To test grammatical comprehension, children were administered the Test of Reception of Grammar (second edition) (TROG: Bishop, 2003). In this assessment, the child hears a sentence and must choose a corresponding picture from a choice of four. Items gradually increase in difficulty, with words becoming less frequent, and structures more complex. The TROG presents numerous structures and syntactic items including spatial prepositions, transitives, subject relative clauses, not only X but also Y sentences (e.g., *the pen is not only small but also blue*), comparatives, passives, coordinated clauses, pronoun forms (plural versus singular), complement clauses and object relative clauses, in approximate order of complexity. It therefore provides a comprehensive measure of receptive grammatical abilities. Children were also administered the British Picture Vocabulary Scale (BPVS: Dunn *et al.*, 2009). This follows the same format as the TROG, with the child hearing a word and being required to choose the correct picture from a choice of four. This assessment was employed to provide a further indicator of the child's language level – specifically, their receptive vocabulary. One child in the L1 did not complete the BPVS, but they were included in the study because the remainder of their data (TROG and collocations task) were complete.

To boost motivation, each child was given a picture of a caterpillar, and at various stages they were allowed to add a sticker to the caterpillar segments. This provided a pictorial record of the children's progress, and was given to the child as a reward at the end of the study.

Ethical approval was obtained from the Newcastle Humanities and Social Sciences Ethics Committee. Informed consent was obtained from the parents, and all children were told that the task was not obligatory, and they could discontinue whenever they wished. Though, due to resource limitations, letters, information sheets and consent forms were in English, both the schools and the research team were available to answer any queries the parents/guardians may have had.

Results

Reliability analysis

The internal consistency of the collocations test was measured using the KR20 statistic (Kuder & Richardson, 1937). This is a version of Cronbach's alpha which is designed for dichotomous data. It produces a value between 0 and 1, with a value > 0.8 generally regarded as acceptable. The KR20 statistic for the full dataset was 0.67. This is close to the level of 0.7 which is often regarded as acceptable in the literature (Taber, 2018). The trimming of items was considered to improve this value. However, to undertake adequate test development, a sufficiently large and homogenous sample is necessary, and this was arguably not the case. Consequently, all items were kept. Nonetheless, where possible, mixed effects models were considered to control for item variability. For the purposes of future test development, item-total correlations are provided in the online materials.

Table 1. Descriptives

Group	n	Age		BPVS raw		TROG blocks		Collocations raw	
		M	sd	M	sd	M	sd	M	sd
L1	40	75.4 (6;3)	4.37	94.8	13.3	12.55	3.33	59.9	6.36
EAL	32	75.4 (6;3)	4.37	74.4	15.6	7.81	3.99	50.7	5.25

Descriptives

The L1 group obtained age-appropriate BPVS and TROG standard scores (m = 99.1, sd = 10.38 and m = 107.2, sd = 14.49 respectively). The EAL group exhibited relatively low scores on these tasks (m = 87.6, sd = 9.22 and m = 85.9, sd = 17.23 respectively). Table 1 reports the means and standard deviations for the children's ages, and their raw scores on the three language tasks.

These data are represented as violin plots in Figure 1. Violin plots have been chosen as a method that visually demonstrates the distribution of the scores and the overlap between groups. The central horizontal line corresponds to the mean. Above the violin plots are the results of tests comparing the groups. While t-tests were conducted for the BPVS, and TROG total scores, a logistic linear mixed effects model was conducted for the collocations task with random by-subject and by-item intercepts. This enabled the modelling of item-level variation. It can be seen that there were significant group differences for all three tasks (all p-values < 0.001). Effect sizes (Cohen's *d*, calculated independently of the statistical models) demonstrate that differences were relatively large, and were largest for the Collocations task (1.53) followed by the BPVS (1.39) and then the TROG (1.28).

For the collocations task, a binomial test was used to determine chance level performance (shown via the grey line). For the L1 group 22.5% of the children performed below chance (n = 9), while for the EAL group, 75% of the group performed below chance (n = 24).

Correlations between language measures

Figure 2 contains a series of scatter plots demonstrating the associations between the different language measures, with a line of best fit. Pearson's correlation coefficients, partialling for age-in-months, are shown above each plot. The variable "Mean z" refers to the mean of the TROG and BPVS z-scores, which were calculated on a group basis. This is intended to provide a measure of general language abilities across linguistic subdomains. It can be seen that in the L1 group, the collocations task was significantly correlated with TROG raw scores (r = 0.51), the BPVS raw scores (r = 0.50), and the mean z-score for these two language measures (r = 0.55). The association between the BPVS and TROG exhibited the strongest correlation (r = 0.68). All of these coefficients are classified as "strong" according to Cohen's (1988) guidelines as they are above 0.50. Correlations were weaker in the EAL group, with only the association between the TROG and BPVS attaining significance (r = 0.58). Other associations were weak to moderate according to Cohen's criteria.
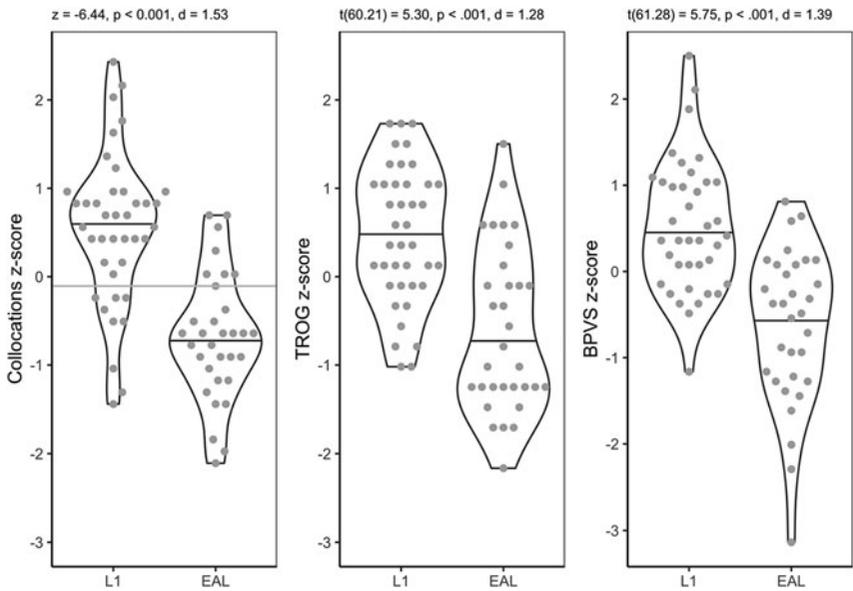
Figure 1. Between-group differences on language tasks
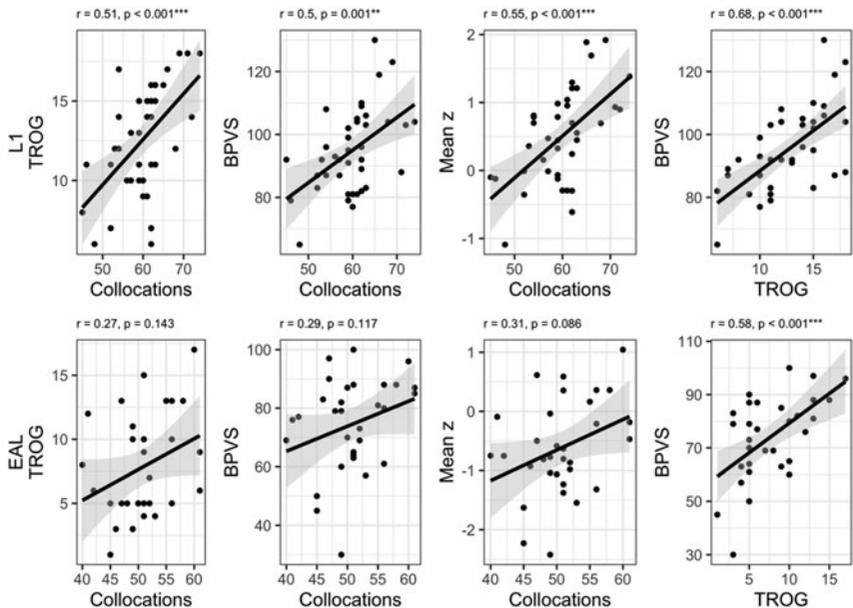Light grey line shows chance performance on collocations task, calculated using a binomial test.

Figure 2. correlations between language measures

Analysis of language profiles

A two-way ANOVA with Greenhouse Geisser correction was conducted to investigate the interaction between Group as between-subjects factor (monolingual versus EAL) and language test as within-subjects factor (TROG versus BPVS versus Collocations Test). The single missing value for the BPVS was replaced with the mean for that group (monolinguals). There was a significant main effect of Group ($F(1, 70) = 53.48$, $p < .001$, $\eta_p^2 = .43$), no significant main effect of Test ($F(2, 140) = 0.01$, $p = .994$, $\eta_p^2 < .01$), and no significant interaction between these two variables ($F(2, 140) = 0.23$, $p = .798$, $\eta_p^2 < .01$). There is therefore no statistical evidence for a qualitative difference in profiles across the two groups.

Discussion

The study had three aims: to determine whether collocational knowledge in children could be assessed with high degrees of reliability and validity; to explore the relationship, in children, between collocational knowledge and other linguistic abilities; and to compare the profiles of monolingual and EAL children. These three aims will be addressed in turn.

Regarding the first aim, the collocations task exhibited desirable psychometric properties. It was significantly associated with other linguistic variables (BPVS and TROG) in the L1 group. These associations were strong according to Cohen's (1988) criteria. This indicates a degree of concurrent validity, defined as a strong association with another measure that is both well-validated, and measures an identical or closely-related construct (Miller, McIntire & Lovler, 2011). Secondly, the collocations task was sensitive to differences between the groups. These were motivated by the assumption, based on Dąbrowska (2019), that the limited exposure of the EAL children would impact on their collocational knowledge. Finally, a measure of internal consistency (KR20) approached acceptability. It is likely that a further study with a larger monolingual group would be able to achieve higher internal consistency by deleting items, whilst simultaneously reducing administration time. To our knowledge the test is only the second assessment of collocational knowledge in children. It targets a younger group than Smith and Murphy's (2015) assessment, and as it uses a mainly visual presentation format it is much less dependent on reading abilities.

Regarding the second question, collocational knowledge was strongly and significantly correlated with other language abilities (receptive vocabulary and grammar) in the L1 group. This is consistent with Dąbrowska *et al.*'s studies of collocational knowledge in adults (Dąbrowska, 2019; Llompart & Dąbrowska, 2020). Such a finding supports the claim that, due to common representational formats across language domains, domain-general learning and memory mechanisms are utilised. However, it is also plausible that these associations arise from domain-specific mechanisms. For example, phonological working memory could be involved in the learning of both single words (Gathercole, 2006), and also multi-word / multi-morpheme units (Ellis & Sinclair, 1996). The latter, in turn, may support the development of both collocational knowledge and morphosyntax. In order to more fully explore the cognitive factors underpinning collocational knowledge, future studies should include more tests of relevant cognitive mechanisms such as phonological short-term memory. Another potentially important mechanism is statistical or procedural learning, though in Llompart and

Dąbrowska ([2020]) the Serial Reaction Time task, often used to assess this construct was not closely association with collocational knowledge.

While correlations were strong and significant across the board in the L1 group, they were weaker in the EAL group with only the correlation between the BPVS and TROG attaining significance. Such differences may have arisen from properties of the measurement scale. For example, the spread of the collocations scores was attenuated in the EAL group, most likely due to floor effects, and this may have weakened the correlation coefficient. It is also possible that differences in correlation coefficients reflect a genuine psychological phenomenon. For example, usage-based accounts have argued that grammatical knowledge may be acquired via analysis of stored chunks (Dąbrowska & Lieven, [2005]; Marchman & Bates, [1994]). However, EAL children may be less dependent on this formulaic route, as they are able to map the structure of the additional language to their already-acquired L1 syntax. There are also societal pressures that may lead to a more analytic learning style in older children, e.g., exposure to print media may encourage a focus on form (Wray & Perkins, [2000]). This more analytic learning style may lead to weaker relationships between language domains, and therefore we would expect to see weaker correlations in the EAL group. Nonetheless, given the issues with measurement scales noted above we should be cautious about making this interpretation. There may be better experimental ways to determine whether EAL children are engaging in a more analytical learning style than their monolingual counterparts, e.g., investigating between-group differences in metalinguistic awareness, or performance on language aptitude tests which depend on analytical skills (e.g., the LLAMA language aptitude test, Meara, [2005]).

It is interesting to note that the relationship between collocational knowledge and vocabulary in the EAL group was substantially weaker in the current study than in Smith and Murphy's study ($r = 0.29$ versus $r = 0.59$). One important difference between the studies is the format of the test. It is possible that, in using an entirely written format, the relationship between collocational knowledge and vocabulary may have been mediated by reading abilities that are relatively weak in this population (Babayiğit, [2014]). Another difference between the studies was the greater age of the children in Smith and Murphy ([2015]), though it is not certain why this should lead to stronger correlations.

Turning to language profiles, the EAL group performed significantly worse across-the-board. Though the effect size for the collocations task was larger than the effect sizes for grammar (TROG) and vocabulary (BPVS), an analysis of the group-by-test interaction did not uncover an uneven profile characterised by greater difficulties on a particular task. This differs from the profile observed by Dąbrowska ([2019]), whereby a significant group-by-test interaction reflected greater between-group differences on the collocations task. A possible explanation for these differences is that the language experience gap, quantifiable as the difference in number of years of regular exposure to English, is much larger in the adult populations studied by Dąbrowska ([2019]) than the child populations on the current study. For example, given that mean age of the monolingual group in the current study was 6;3, this places an effective upper limit on the experience gap which may be greatly exceeded among adults. This may result in greater group differences among adults on tasks that are, putatively, sensitive to exposure, e.g., the collocations task.

In terms of understanding language development in EAL children, the findings of the current study support Smith and Murphy's ([2015]) finding that collocational

knowledge for English is impaired. This may impact on other aspects of educational performance. Smith and Murphy argue that weak collocational knowledge may impact on reading decoding and fluency – for example, resulting in a word-by-word analysis of idiomatic expressions. Reading is a key area of difficulty for EAL children. For example, while, in terms of overall academic performance, EAL children catch up with monolingual peers by late adolescence, their reading is often weaker than other academic skills, e.g., mathematics (Strand, Malmberg & Hall, 2015). Another potential area of difficulty is word-learning. Breadth of lexical knowledge may be diminished compared to L1 peers even in early adolescence (ages 13–14) after a decade of formal schooling (Cameron, 2002). This may, in part, reflect difficulties with collocations, given that the collocational properties of words may act as important cues to meaning (Dabrowska, 2009). Such lexical limitations may also drive reading difficulties. For example, Babayiğit (2014) finds that lexical knowledge is an especially strong predictor of reading skills in EAL versus monolingual children. Given the potential interplay between collocational and lexical knowledge, and the potential influence of lexical knowledge on reading abilities, there is emerging evidence that it would be beneficial for school teachers to help children with EAL to develop better collocational knowledge.

However, research in this area is at an early stage, so it would be premature to advocate an extensive programme of intervention in this area, until the impact of limited collocational knowledge has been fully elaborated. In addition, given the heterogeneity among the EAL population, we must be careful about making strong generalisations. In a large-scale analysis of the UK National Student Database, Strand *et al.* (2015) argues that "the heterogeneity within the EAL group is so large that the average EAL [- English as a First Language] gap is fairly meaningless in comparison" (p. 7). Other factors which impact on educational performance are ethnicity, deprivation, and timing of arrival in the UK education system. Moreover, EAL status may even incur benefits. For example, Whiteside, Gooch and Norbury (2017) found that children with EAL (Reception and Year 2) demonstrated better social emotional and behavioural development, and were more likely to meet academic targets, than L1 peers matched on English abilities. Moreover, research has also indicated that, even within apparently homogenous EAL groups, subsamples may be identified. For example, within a large group of EAL children with an L1 Spanish-speaking background, Kapantzoglou, Restrepo, Gray and Thompson, (2015) were able to identify three subgroups based on L1 measures: one with grammatical difficulties, one with working memory difficulties, and a third with average skills across-the-board.

Future research on the topic of collocational knowledge is necessary to support the claims of the current study. By running the existing task with a larger homogenous monolingual study, it may be possible to create a test with fewer items, a shorter administration time, and higher internal consistency. Further cognitive variables may be assessed to determine, and covary for, alternative sources of individual differences, e.g., tests of working memory, or statistical learning abilities (e.g., Kidd & Arciuli, 2015). The inclusion of a larger battery would also further clarify the claim that relationships between linguistic variables reflect domain-general processes (Dąbrowska, 2019; Llompart & Dąbrowska, 2020). Investigation of the L1 backgrounds of the EAL children may also enable statistical control over variation in general language abilities. This is a variable that could not be controlled for in the current study. To achieve this, a more detailed measure of exposure (e.g., Cattani *et al.*, 2014) would help to

determine whether collocational difficulties are driven mainly by limited language experience. Longitudinal modelling may help to determine whether relationships are genuinely causal. Finally, if there is sufficient evidence that collocational knowledge is both weak in children with EAL, and impacts on other language domains and academic performance, there would be a strong incentive to develop an educational intervention in this area.

References

Adams, C., Cooke, R., Crutchley, A., Hesketh, A., & Reeves, D. (2001). *Assessment of Comprehension and Expression 6–11 (ACE 6–11)*. NFER-Nelson.

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

Benor, S. B., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language*, 233–278.

Bishop, D. V. M. (2003). *Test of Reception of Grammar (Trog-2)*. Harcourt Assessment.

British National Corpus, Version 2. (2002). BNC Consortium.

Bybee, J. L. (2010). *Language, usage and cognition* (Vol. 98). Cambridge University Press Cambridge.

Cain, K., Towse, A. S., & Knight, R. S. (2009). The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, *102*(3), 280–298. https://doi.org/10.1016/j.jecp.2008.08.001

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research*, *6*(2), 145–173. https://doi.org/10.1191/1362168802lr103oa

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500–508.

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests?: Language exposure and bilingual screening. *International Journal of Language & Communication Disorders*, *49*(6), 649–671. https://doi.org/10.1111/1460-6984.12082

Chomsky, N. (1957). *Syntactic Structures*. Mouton.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

Culicover, P. W., Jackendoff, R., & Audring, J. (2017). Multiword Constructions in the Grammar. *Topics in Cognitive Science*, *9*(3), 552–568. https://doi.org/10.1111/tops.12255

Dabrowska, E. (2009). Words as constructions. *New Directions in Cognitive Linguistics*, 201–224.

Dąbrowska, E. (2014). Words that go together: Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon*, *9*(3), 401–418.

Dąbrowska, E. (2019). Experience, Aptitude, and Individual Differences in Linguistic Attainment: A Comparison of Native and Nonnative Speakers. *Language Learning*, *69*(S1), 72–100. https://doi.org/10.1111/lang.12323

Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*(3), 437–474. https://doi.org/10.1515/cogl.2005.16.3.437

Davies, M. (2004). BYU-BNC. (Based on the British National Corpus from Oxford University Press). http://corpus.byu.edu/bnc/

Dunn, L. M., Dunn, L., & Styles, B. (2009). *The British Picture Vocabulary Scale—3rd Edition*. GL-Assessment.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. *The Handbook of Second Language Acquisition*, 63–103.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocbulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology*, *1*(49A), 234–250.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(04), 513–543.

Gibbs, R. W. (Jr). (2010). Idioms and Formulaic Language. *The Oxford Handbook of Cognitive Linguistics*. https://doi.org/10.1093/oxfordhb/9780199738632.013.0027

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. University of Chicago Press.

Kapantzoglou, M., Restrepo, M. A., Gray, S., & Thompson, M. S. (2015). Language Ability Groups in Bilingual Children: A Latent Profile Analysis. *Journal of Speech, Language, and Hearing Research*, *58* (5), 1549–1562. https://doi.org/10.1044/2015_JSLHR-L-14-0290

Kidd, E., & Arciuli, J. (2015). Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Development*, n/a-n/a. https://doi.org/10.1111/cdev.12461

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2* (3), 151–160. https://doi.org/10.1007/BF02288391

Langacker, R. W. (2009). Cognitive (construction) grammar. *Cognitive Linguistics*, *20*(1), 167–176.

Llompart, M., & Dąbrowska, E. (2020). Explicit but Not Implicit Memory Predicts Ultimate Attainment in the Native Language. *Frontiers in Psychology*, 11, 569586. https://doi.org/10.3389/fpsyg.2020.569586

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, *21*(02). https://doi.org/10.1017/S0305000900009302

Meara, P. (2005). LLAMA language aptitude tests: The manual. *Swansea: Lognostics*.

Miller, L. A., McIntire, S. A., & Lovler, R. L. (2011). *Foundations of Psychological Testing: A Practical Approach*. SAGE.

Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, *24*(2), 223–242. https://doi.org/10.1093/applin/24.2.223

Nippold, M. A., & Martin, S. T. (1989). Idiom interpretation in isolation versus context: A developmental study with adolescents. *Journal of Speech, Language, and Hearing Research*, *32*(1), 59–66.

Pine, J., & Lieven, E. (1997). Slot and frame patterns in the development of the English determiner category. *Applied Psycholinguistics*, *18*, 123–138.

Schutze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. University of Chicago Press.

Smith, S. A., & Murphy, V. A. (2015). Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing*, *28*, 347–369. https://doi.org/10.1007/s11145-014-9527-y

Strand, S., Malmberg, L., & Hall, J. (2015). English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard UP.

Verhagen, V. (2019). *Illuminating variation: Individual differences in entrenchment of multi-word units*.

Whiteside, K. E., Gooch, D., & Norbury, C. F. (2017). English Language Proficiency and Early School Attainment Among Children Learning English as an Additional Language. *Child Development*, *88*(3), 812–827. https://doi.org/10.1111/cdev.12615

Wray, A. (1998). Protolanguage as a holistic system for social. *Language and Communication*, *18*, 47–67.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463–489. https://doi.org/10.1093/applin/21.4.463

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, *20*(1), 1–28.

Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The growth of language: Universal Grammar, experience, and principles of computation. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2016.12.023

Search Strings for Corpus Search

Search strings were written in the "Corpus Query Language". Here are some examples:

1. **Binomials, e.g. *apples and pairs***

[lemma = "apple"] [word !== "." & word !=="?" & word !=="!"]{0,2} [lemma = "pear"] within <s/>

Note that the search string rules out the inclusion of sentence-final delimiters (".", "?","!") to ensure that strings do not straddle sentence boundaries. This is also ensured by the "within <s/>" statement. Lemmas are searched for, which allows for pluralisation of singular nouns, e.g. *pair/pairs*, and inflection of verbs

For *fruit and vegetables*, the abbreviated form *veg*, was allowed at the end of the search string. For *stop and go*, a present participle was not allowed at the end of string to rule out *stop going*, in which the two verbs do not have equivalent lexical status.

2. **Other collocations, e.g. *make a mess, as black as ink*.**

[lemma = "make"] [word !== "." & word !=="?" & word !=="!"]{0,2} [lemma = "mess"] within </s>

[lemma = "black"] [word !== "." & word !=="?" & word !=="!"]{0,2} [lemma = "ink"] within </s>