

Appropriateness as an aspect of lexical richness: What do quantitative measures tell us about children's writing?

Philip Durrant^{a,*}, Ayça Durrant^b

^a University of Exeter, St Luke's Campus, Heavitree Road, Exeter EX1 2LU, UK

^b INTO, University of Exeter, Streatham Campus, Stocker Road, Exeter EX4 4PY, UK

ARTICLE INFO

Keywords:

Vocabulary
Children's writing
Academic vocabulary
Register
Corpus linguistics

ABSTRACT

Quantitative measures of vocabulary use have added much to our understanding of first and second language writing development. This paper argues for measures of register appropriateness as a useful addition to these tools. Developing an idea proposed by [Durrant and Brenchley \(2019\)](#), it explores what such measures can tell us about vocabulary development in the L1 writing of school children in England and critically examines how results should be interpreted. It shows that significant patterns of discipline- and genre-specific vocabulary development can be identified for measures related to four distinct registers, though the strongest patterns are found for vocabulary associated with fiction and academic writing. Follow-up analyses showed that changes across year groups were primarily driven, not by the nature of individual words, but by the overall quantitative distribution of register-specific vocabulary, suggesting that the traditional distinction between measures of lexical diversity and lexical sophistication may not be helpful for understanding development in this context. Closer analysis of academic vocabulary showed development of distinct vocabularies in Science and English writing in response to sharply differing communicative needs in those disciplines, suggesting that development in children's academic vocabulary should not be seen as a single coherent process.

1. Introduction

Researchers and educators have long been interested in assessing and modeling writing development with quantitative measures of language use (e.g., [Crossley, 2020](#); [Durrant, Brenchley, & McCallum, 2021](#); [Leńko-Szymańska, 2020](#)). Such research typically aims to provide objective linguistic measures by counting the frequency with which particular features occur, or by quantifying the level of sophistication or complexity of those features. These measures are then used to track differences across developmental time, between learners at different levels of proficiency, or between texts that have been graded at different levels of quality.

This approach to writing assessment has two key strengths. First, because analyses can be partly automatized through easy-to-use NLP applications (e.g. [Kyle & Crossley, 2015](#); [Lu, 2010](#)), it allows for analysis of large numbers of texts. This can enable robust generalizations, highlight subtle patterns that are difficult to spot in smaller samples, and show how development differs across learner groups, text types and contexts. Secondly, quantifying language use requires researchers to be specific about the features they are studying and explicit about how those features are defined and identified. This can promote the rigor and transparency of research and provide findings that are detailed and specific.

* Corresponding author.

E-mail addresses: p.l.durrant@exeter.ac.uk (P. Durrant), a.durrant2@exeter.ac.uk (A. Durrant).

A prominent line of research in this tradition has focused on development in vocabulary (Kyle, 2021; Leńko-Szymańka, 2020). This is an especially important focus in the context of child writing as children's vocabulary is known to be a key predictor of academic success (e.g. Durham, Farkas, Hammer, Tomblin, & Catts, 2007; Spencer, Clegg, Stackhouse, & Rush, 2017; Townsend, Filippini, Collins, & Biancarosa, 2012) and to vary strikingly across children from different socio-economic groups and with different linguistic backgrounds (e.g. Collins & Toppelberg, 2021; Meir & Armon-Lotem, 2017; Spencer, Clegg, & Stackhouse, 2012). This implies a pressing need both for clear models of vocabulary development and for reliable and easy-to-use means of assessing such development (Roessingh, Douglas, & Wojtalewicz, 2016).

Most vocabulary-oriented work has focused on the *diversity* of words that writers use and/or the *sophistication* of those words. Vocabulary sophistication has been influentially defined as "selection of low-frequency words that are appropriate to the topic and style of writing, rather than just general, everyday vocabulary" (Read, 2000, p. 200). While this definition points to two aspects of sophistication (low-frequency and appropriateness to topic/style) most research has focused only on the former (Leńko-Szymańka, 2020). This is understandable; frequency is relatively easy to quantify and many studies have found it to correlate with development over time or with ratings of quality (e.g., Olinghouse & Leaird, 2009; Roessingh, Elgie, & Kover, 2015; Sun, Zhang, & Scardamalia, 2010). However, this focus on frequency has left the important second part of Read's definition out of the developmental picture. That is, there has been little interest in the issue of appropriateness to topic/style of writing.

We take appropriateness to be essentially an issue of register: a text is more sophisticated if the writer selects words that are characteristic of the register they are attempting to invoke. The concept of *register* refers to the way in which situational contexts and the writer/reader purposes that accompany them give rise to characteristic sets of linguistic features (Biber & Conrad, 2009). Thus, for example, the context of a television news broadcast differs in characteristic ways from that of a telephone call between friends. The former involves one-way communication, visual images, no personal relationship between speaker and audience, and an informational focus, often surrounding themes such as politics and sport. The latter typically involves two-way communication, no visual images, a personal relationship between interlocutors, and often a phatic, rather than informational focus. Situational differences of this sort are associated with differences in language use as interlocutors orient towards the functional needs and social conventions of the situation.

As Biber and Conrad (2009) note, registers can be defined in multiple ways and at multiple levels of granularity: *academic prose* is a very broad register; medical research textbooks is more specific. The level of granularity at which registers are defined ideally depends on the aims of the classification and knowledge about the extent and nature of contextual variation. In practice, it is also influenced by the practical issue of what resources are available to the researcher.

The importance of establishing measures of appropriateness has been noted by Leńko-Szymańka (2020), who found this to be a crucial part of the rationale given by raters for vocabulary grades awarded to texts. Raters further noted that supposedly "advanced" words can sound "forced and contrived" unless they fit the register the writer is attempting to invoke (2020, p. 222).

The potential contribution of a measure of appropriateness to our understanding of vocabulary development can also be theorized in terms of the distinction between *breadth* and *depth* of vocabulary knowledge, whereby the former identifies what words learners know and the latter identifies what they know about those words (e.g., Schmitt, 2014). Most sophistication measures are primarily measures of vocabulary breadth; they tell us something about the repertoire of words that learners know, rather than what they know about those words. Indeed, the mere ability to write a word down in a text (regardless of that word's semantic accuracy, register appropriateness, or integration into its syntactic or collocational context) suffices for that word to be notched up as part of the writer's repertoire. A measure of register appropriateness could introduce a stronger element of vocabulary depth by determining whether a learner demonstrates awareness of the contexts in which a word should be used.

Previous research has addressed register tangentially in terms of *academic* (Roessingh et al., 2016; Sun et al., 2010; White, 2015) or *Greco-Latin vocabulary* (Berman & Nir-Sagiv, 2007; Corson, 1985, 1989), both of which have been found to increase in prominence as children mature. However, this line of work is limited in that it considers development only in terms of a single erudite register and in that it makes no attempt to relate this register to the types of text that children are being asked to write. This second point is crucial as use of academic vocabulary is only *appropriate* when children are attempting to invoke an academic register. In other contexts, academic vocabulary may be jarringly inappropriate, and therefore unsophisticated.

To address these limitations, Durrant and Brenchley (2019) proposed a measure which evaluates separately the extent to which a text's vocabulary is typical of different registers (see Section 2.2 for details). Focusing on the contrasting registers of academic and fiction writing, they find that children's stories use fiction-like vocabulary extensively from the youngest ages and maintain this throughout their education. Their non-fiction writing, in contrast, shows a steady drop in fiction-like vocabulary and a substantial increase in academic vocabulary, demonstrating a clear shift towards a more mature academic style. Stories also see a small increase in academic vocabulary, suggesting that their maturing academic style is carrying over into their fiction writing, but this increase is modest compared to that seen for non-fiction.

While these findings are promising, much remains unclear. One limitation is that Durrant and Brenchley considered development in terms of only two registers: fiction and academic. A second is that it did not distinguish between learner writing in different academic disciplines. This is problematic because written vocabulary is known to differ dramatically across disciplines (Durrant, 2014, 2016; Hyland & Tse, 2007). Any model that attempts to capture children's vocabulary development at school therefore needs to take account of such variation.

A more fundamental issue with the proposed measure is that it is not clear how differences in scores should be interpreted. A plausible reading of an increased score for academic vocabulary use, for example, is that it shows writers developing a vocabulary that better matches the norms of the target register. This would be a helpful outcome, implying that the measure tells us something specific about development in writing proficiency. However, there are reasons for questioning this interpretation.

One issue is that the words children write are not pure reflections of their vocabulary proficiency. Writing is, as has been widely discussed in the literature (e.g. Grabe & Kaplan, 1996; Hayes, 2012; Shaw & Weir, 2007), a highly complex product. The words that end up on the page are the result, not only of the writer's vocabulary knowledge but of, amongst other things, the topic on which they are writing, their intended audience, the resources available to them (dictionaries, computers, teacher input, etc), their personal goals and motivations, and so on. We therefore need to exercise caution in ascribing reasons for any quantitative developmental patterns. While these *may* reflect developments in particular aspects of writer proficiency (e.g., vocabulary knowledge), they may equally reflect shifts in the topics discussed, in the resources available, or any of a range of other variables. Careful inspection of the individual cases that underlie the general patterns is therefore needed before conclusions can be drawn.

A second issue is that the measure of appropriateness is based on the frequency distributions of words within a specific reference corpus (see Section 2.2). The validity of the measure therefore depends on the assumption that the frequencies attested in that corpus are representative of those in the target registers more broadly.

Representativeness is a problematic concept in corpus linguistics (e.g. Leech, 2007). To take academic corpora as our example once again, it is not obvious what range of disciplinary areas and what text genres need to be included in a representative corpus or how they should be balanced in terms of their relative sizes. It is not obvious if all disciplines should be represented with a similar number of texts, or if disciplines that produce more publications (e.g. Medicine) or that teach a larger number of students (e.g. Business) should be given greater weight than more niche concerns. Indeed, it is not even obvious what should count as a discipline. Should, for example, Biomedical Science, Nursing and Medical Imaging be treated as distinct disciplines or classified under the single heading of Medicine and Health?

There is no universally correct answer to these questions and decisions need to be made according to the aims of individual projects. However, we are left with the concern that frequency data from a reference corpus may skew the measure of appropriateness in misleading ways. To take an extreme example, an academic corpus based entirely on medical research articles would not yield a satisfactory measure of vocabulary appropriateness for children's History essays.

Though the COCA corpus, on which Durrant and Brenchley's register measure was based, is large and wide-ranging (Davies, 2008), hidden biases related to topics, genres, sources, etc. may skew overall frequencies towards or away from particular words. Indeed, in the academic case there is reason to believe that biases do exist, since a vocabulary list built on the corpus has been shown to be skewed across academic disciplines (Durrant, 2016). Such concerns again imply that broad quantitative patterns cannot be taken at face-value. Again, therefore, to understand what those patterns mean, it is important to inspect the individual cases that underlie them.

The present study will investigate these issues with the joint aims of increasing our understanding of register appropriateness as a measure of lexical sophistication, and of learning more about the development of vocabulary use in children's writing. Specifically, it addresses two questions:

1. How does the register appropriateness of children's written vocabulary develop in quantitative terms as they progress through their education?
2. How should developmentally significant patterns of register appropriateness be interpreted?

2. Methods

2.1. Corpus

The data for this study came from the *Growth in Grammar* corpus¹, a collection of almost 3000 texts collected between 2015 and 2017 from approximately 1000 children in 24 schools across England. Texts were written as part of children's regular schoolwork and were collected and digitized with the permission of children, their parents/guardians, teachers, and participating schools. Approximately 13% of texts were written by children having English as an Additional Language (slightly below the national average at the time) and 22% were written by children in receipt of pupil premium/free school meals (slightly above the national average; Department for Education, 2018). Texts were classified as *literary* or *non-literary* in genre based on their overall purpose (as detailed in Durrant & Brenchley, 2019). The former most characteristically included works of creative fiction or literary imitations. The latter included historical recounts, literary criticism, experimental reports, and persuasive speeches. Texts were sampled from three disciplinary areas: English, Humanities (History, Geography, and Religious Studies), and Science.

Texts were predominantly handwritten. These were typed and checked by a small team of transcribers, who removed any material (including personal, institutional, or place names and any relevant details of writers' own lives or circumstances) that might compromise the anonymity of children, participating schools, or other connected individuals.

For this study, we sampled texts written by primary school children in Years 2 and 6, and by secondary school children in Years 9 and 11. These groups correspond to the endpoints of the four Key Stages of the English education system. Table 1 summarizes the

¹ Full details and information about access can be found at: <http://reshare.ukdataservice.ac.uk/853809>

Table 1
Corpus makeup.

		Number of				Words per text ^a		
		Texts	Schools	Writers	Titles	Mean	Min	Max
Year 2	English Literary	258	5	95	43	66	4	333
	English Non-Literary	278	6	149	24	53	3	344
	Humanities	96	4	87	7	53	6	283
	Science	7	2	6	2	37	16	68
Year 6	English Literary	293	6	177	32	318	39	1019
	English Non-Literary	298	7	142	23	150	6	517
	Humanities	106	4	80	12	163	52	688
	Science	171	4	90	11	165	39	514
Year 9	English Literary	220	4	220	7	341	69	2030
	English Non-Literary	305	7	165	29	270	35	994
	Humanities	113	3	40	17	216	55	621
	Science	166	3	74	33	83	4	1011
Year 11	English Literary	63	4	45	8	345	128	1188
	English Non-Literary	368	6	110	60	319	55	1423
	Humanities	49	1	19	8	214	75	876
	Science	58	2	38	14	218	52	2054
	Overall	2849	24	969	327	235	3	2054

^a 'words' are defined as tokens that were attested in COCA. Any items not attested in the reference corpus were excluded from the analysis and from the word count.

Table 2
Calculating register score for the noun *analysis*.

	Academic	Fiction	Magazine	News	Summed frequency
Normed frequency	561.11	12.87	54.32	47.13	561.11 + 12.87 + 54.32 + 47.13 =
Register score	561.11/675.43 = 0.83	12.87/675.43 = 0.02	54.32/675.43 = .08	47.13/675.43 = .07	675.43

makeup of the corpus. Because few Year 2 Science texts were available (reflecting a lack of sustained Science writing at this level), these were excluded from the analysis (Table 2).

2.2. Analysis

The measure of register appropriateness introduced in Durrant and Brenchley (2019) aims to determine how characteristic a text's vocabulary is of a given target register. This raises the question of how a target register should be defined. In this regard, a distinction can be made between a *peer* target and a *prospective* target. The former evaluates writing in relation to successful writing by learners' contemporaries; the latter evaluates writing in relation to writing which learners might reasonably aim to produce at a later stage of development. Peer targets are useful because the types of writing that learners produce change throughout their education. Because different types of writing require different types of language, text produced by more 'advanced' writers may not always be an ideal model (Hyland, 2008). Peer models are therefore a good way of evaluating how learners' writing relates to the quality expectations that are most relevant to them. However, this also implies that a different measure of appropriateness is needed for each group, so preventing any comparison across stages. To understand how writing develops as learners advance through their education, we need a single measure that can be applied across the course of this education. For purposes of tracing development over time, therefore, target registers are better defined prospectively.

A second issue is the specificity of the reference corpus. That is, we need to decide the granularity of the register with which learner writing should be compared. As we have seen, registers can be relatively broad (e.g., *academic*) or relatively specific (e.g., *medical research textbooks*). In the current research, reference corpus registers are defined at a broad level for three key reasons. First is the practical point that large fine-grained reference corpora are not readily available and so not a viable tool for writing assessment. Second, it is not obvious how fine-grained reference corpora should be defined. This would require a clear understanding of the types of

variation that occur within a register (e.g., between different types of magazines) and of how this variation relates to curricular goals. In the absence of such an understanding across the four major registers studied, a more coarse-grained approach is preferred. Third, it is not clear that more fine-grained registers would be equally applicable across year groups. For example, while children in all year groups wrote stories, the genre of these stories evolved, with the youngest children's stories often focusing on fantastical themes and recalling the style of fairy tales, while older children's writing was often influenced by dystopic fiction or the natural environment.

With these points in mind, our measure of appropriateness is based on the frequency lists created by Davies (2018), which show the frequency per million words of 100,000 word forms in each of the five registers in the *Corpus of Contemporary American English* (Davies, 2008): *spoken, academic, fiction, magazine* and *newspaper*. Because the spoken subcorpus does not represent a clear and coherent register (including a wide range of formal and informal, scripted and unscripted spoken language), our analysis focused on the four written registers only.

These frequency counts were lemmatized (to create a list of approximately 70,000 lemmas) and converted into *register scores*. Register scores represented the relative proportion of normalized occurrences of each word found in each register. As exemplified in Fig. 1, this was calculated by summing the normalized frequencies of the word across the four subcorpora then separately dividing the frequency for each register by that total. This yielded four numbers, summing to 1, which represented the relative proportion of occurrences found in each register (as illustrated in Table 3). Words with register scores of close to .25 in all four registers (e.g. *the*) were evenly distributed, while scores above (or below) .25 in a given register showed that the word was skewed towards (or away from) that register (e.g. *analysis*, which is skewed towards the academic register and away from all other registers, and *happy*, which is skewed towards the fiction register and away from the academic register – as shown in Table 3).

To prepare the study corpus for analysis, it was first tagged for part of speech using CLAWS (Garside & Smith, 1997). The resulting tags were then simplified using a search-and-replace script² to align the tagset with that used in the COCA lists. British spellings were converted to US using the list available at <http://www.tysto.com/uk-us-spelling-list.html>³ ⁴. Frequency data for each lexical word (adjective, adverb, noun and lexical verb) in each text was retrieved from the COCA lists and their register scores calculated. Words in these counts were defined at the *lemma* level. That is, differently inflected forms of a word (e.g., *argue, argues, argued, arguing*) were treated as the same item, but words with the same spelling but different parts of speech (e.g., *mean* as verb, adjective, and noun) were distinguished. Each text was then assigned an overall score for each of the four registers, calculated as the mean score for that register of all lexical words in the text.

Our first aim is to determine how register measures vary across year groups and text types. We will therefore conduct four main statistical analyses: one for each register measure, in which the measures are the dependent variables and year group and text type are the main independent variables. Text types were defined in the first place in terms of academic discipline: English, Humanities and Science. Because writing within the English discipline included both literary and non-literary writing, which differ markedly in their use of register-specific vocabulary (Durrant & Brenchley, 2019), these were further divided into two separate groups.

As in most learner corpus studies, texts were not independent data points: some children contributed multiple texts; individual schools and individual assignment tasks were represented by multiple children. This required the use of mixed-effects models (Gries, 2015), which were implemented using the *lme4* (Bates, Mächler, Bolker, & Walker, 2015) package in *R*. In these models, year group, academic discipline and text genre were fixed effects. School, child (nested within school) and assignment task were included as random effects. Following Winter (2019), the significance of each fixed effect was evaluated using a likelihood ratio test to compare the full model with the model after the effect had been removed. Marginal (excluding random effects) and conditional (including random effects) model fits were calculated using the *r.squaredGLMM* function within the *MuMIn* package (Barton, 2020). To check that data met the assumptions required for accurate and generalizable mixed-effect modules (Tabachnick & Fidell, 2014), histograms and QQ-plots of residuals were checked for outliers and normal distribution; plots of residuals vs. observed values were checked to confirm linearity; plots of standardized residuals versus fitted values were checked to confirm homoscedacity of residuals. All analyses met the assumptions.

3. Findings

3.1. Distribution of register score across disciplines/genres and year groups

Fig. 1 shows the variation in scores for each register across disciplines/genres and year groups. Table 4 summarizes the mixed effects models for each register.

Several points can be taken from these data. First, the fixed effects of discipline/genre and year group are significant for all register measures. Interactions between these effects are also significant for all measures except magazine. The model fits are strongest for the academic and fiction measures (where marginal R^2 s are .59 and .61 respectively) but fits for magazine and news measures are also respectable (with marginal R^2 s of .38 and .26). Random effects were strong (indicated by conditional R^2 s being substantially higher than marginal R^2 s), especially those for topic, suggesting that individual assignments have a strong influence on register-specific vocabulary.

The plots in Fig. 1 enable us to further interpret these models. These show that:

² The analyses reported throughout this paper were carried out using *R* (R Core Team, 2014).

³ Last accessed 22nd March 2021.

⁴ This list was amended to include the word *mum*, which was changed to US *mom*.

- *Academic vocabulary* is used most intensively in science writing and least in literary English writing. Use of academic vocabulary increases across year groups for all text types but the increase in literary English writing is much less than that in other text types.
- *Fiction vocabulary* is used most intensively in literary English writing and least in Science and Humanities writing. Use remains constant across year groups in literary writing but decreases in all other text types.
- *Magazine vocabulary* is used most intensively in Science writing, but its use decreases across year groups. Other text types are similar to each other and use is lower in secondary than primary school.
- *News vocabulary* is not prominent, all mean scores being below .20. Despite this, it does show clear discipline/genre-dependent variation, being most intensively used in Humanities writing and least in Science writing. Only Humanities writing shows a development trend, with use increasing across year groups

These data show that all four measures of register-specific vocabulary are significant markers of both discipline/genre and year group in children's writing. As we argued in Section 1 above, however, questions of interpretation remain. That is, we need to understand *why* these measures vary in the ways they do. A separate investigation of each of the patterns described above is beyond what can be reported within a single article so the following sections will focus on two key sets of questions:

1. Magazine-type words

- a. Why are magazine-type words so prevalent in Science writing?
- b. Why does this prevalence drop so markedly across year groups?

2. Academic words

- a. Why does the academic score increase across year groups?
- b. Given that such vocabulary is known to differ markedly between academic disciplines, does this increase reflect similar developments across subject areas, or do different areas have different profiles of development?

Addressing these questions in detail can both provide insight into children's writing development and, from a broader methodological perspective, throw light on the types of effects that underlie measures of appropriate vocabulary.

3.2. The use of magazine-like words in Science writing

The approach taken in this section is to identify the words in children's Science writing that are particularly magazine-like and then to examine the nature and distribution of those words. *Magazine-like* vocabulary is defined as words that have a magazine register score more than one standard deviation above the mean for COCA. On this definition, 14% of word types in the COCA list were classified as magazine-like. Of these, 211 distinct types were found in the GiG science texts. The full list of types found for each year group is shown in Appendix A.

A glance at these lists reveals that many magazine-like words look stereotypically 'scientific' in nature. To corroborate this impression, each word was checked for its appearance in the Oxford Reference *Science and Technology* series⁵. This is a searchable set of 61 dictionaries and encyclopedia devoted to various fields of science and technology (e.g. *Dictionary of Chemical Engineering; Dictionary of Genetics; Berkshire Encyclopedia of Sustainability*) which represents an authoritative and relatively comprehensive guide to scientific vocabulary. Any word appearing either as a main entry or within a definition in any of the books within this series was deemed appropriate to scientific writing.

Although a small number of words did not appear, with the greatest number of non-scientific words in Year 6 texts (e.g., *power-packed; scrumptious; awesomely*), most did, scientific words comprising close to 99% of tokens at Year 6, and almost 100% for Year 11 (see Table 5). The prevalence of magazine-like words in these texts does not, therefore, imply that children use large numbers of words that are inappropriate to science writing. It seems, rather, to indicate that many words that are appropriate to scientific writing score highly on the magazine measure; suggesting in turn that the magazines featured in COCA make prominent use of scientific vocabulary.

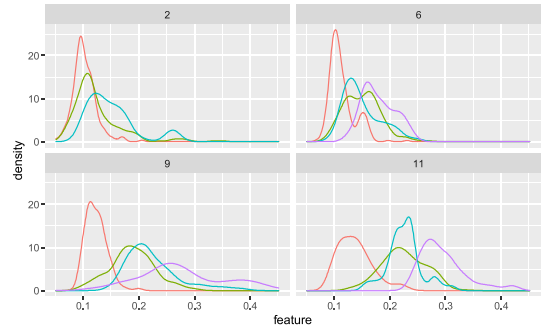
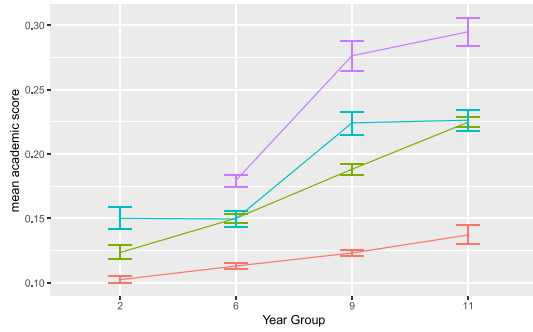
It is important to remember that not all magazine-like words contribute equally to the quantitative patterns that we are trying to explain. Words increase the overall magazine score as a function of two things: their frequency (words which are repeated more often play a larger role in the overall score) and their magazine score (words which are more distinctively magazine-like increase the overall score more than those which are less distinctive). To account for this, a measure of *magazine weight* was calculated for each magazine-like word by multiplying the word's normalized frequency by its magazine score. This gives a rough-and-ready measure of the extent to which each word contributes to the overall patterns shown in Fig. 1.

Fig. 2 shows the distribution of magazine scores for each year group. As this figure demonstrates, magazine weights were strongly skewed, implying that a small number of types accounted for the majority of the overall magazine scores reported in Fig. 1. This was especially the case at Year 6, where just 20 types accounted for 85% of the total magazine weight of the 107 magazine-like words. At Years 9 and 11, the top 20 types accounted for 80% and 74% of the total respectively. We therefore see a tendency for the skew towards a small number of influential words to decrease across the year groups.

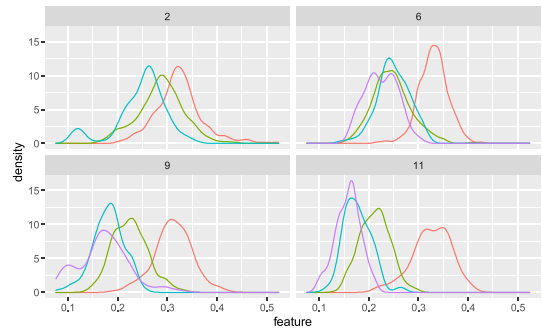
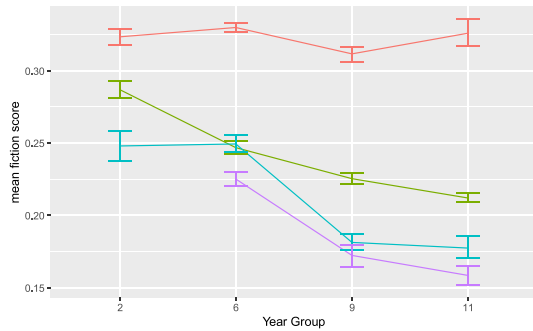
Because the top 20 words by weight account for a large proportion of the overall magazine weight, inspection of these words should

⁵ Retrieved from <https://www.oxfordreference.com> on 3rd February 2020.

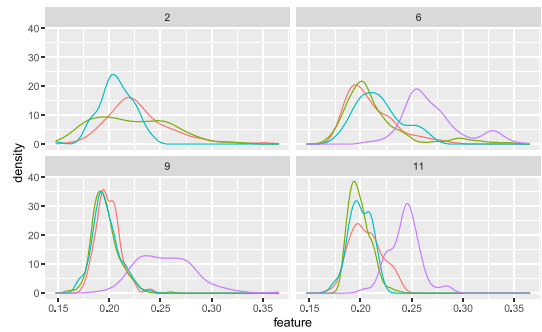
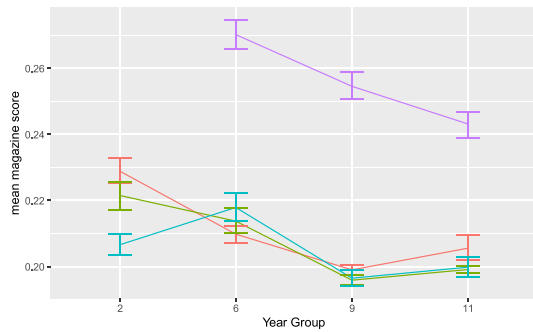
Academic



Fiction



Magazine



News

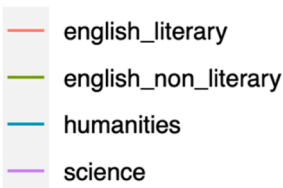
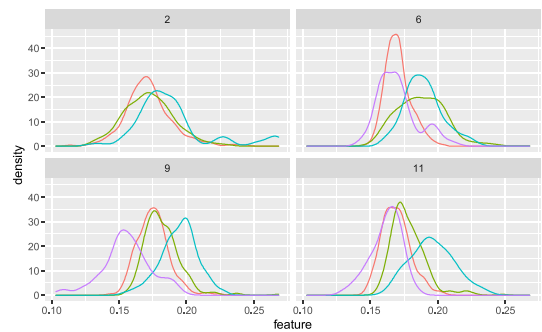
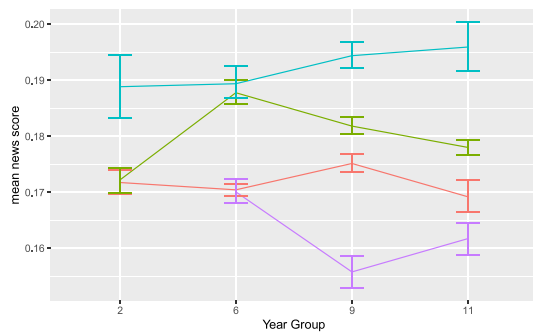


Fig. 1. Distribution of register measures across year group and disciplines/genre.

Table 3

Example register scores.

Lemma	POS	Register score			
		Academic	Fiction	Magazine	News
the	determiner	0.28	0.24	0.24	0.24
analysis	noun	0.83	0.02	0.08	0.07
happy	adjective	0.07	0.41	0.28	0.24

Table 4

Summary of mixed effects models.

Fixed effects*: χ^2			Random effects			R ²		
Discipline/ genre	Year group	Interaction	Effect	Variance	Std.Dev	Marginal	Conditional	
Academic	147.16**	81.25**	51.11**	writer in school	9.19e ⁻⁵	.01	.59	.88
				school	1.48e ⁻⁴	.01		
				topic	9.99e ⁻⁴	.03		
				residual	5.06e ⁻⁴	.02		
Fiction	250.27**	82.42**	57.55**	writer in school	6.63e ⁻⁵	.01	.61	.85
				school	8.88e ⁻⁵	.01		
				topic	8.63e ⁻⁴	.03		
				residual	6.57e ⁻⁴	.03		
Magazine	107.97**	38.56**	13.17	writer in school	.00	.00	.38	.77
				school	.00	.00		
				topic	3.74e ⁻⁴	.02		
				residual	2.16e ⁻⁴	.01		
News	96.82**	38.22**	23.58*	writer in school	5.67e ⁻⁶	.00	.26	.65
				school	7.29e ⁻⁶	.00		
				topic	1.19e ⁻⁴	.01		
				residual	1.19e ⁻⁴	.01		

* $p < .005$.** $p < .0001$; df for main effects = 11; df for interactions = 8.

give us a good idea of the factors influencing the overall patterns we are trying to explain. These words are reproduced in Table 6 and summarized quantitatively in Table 7. The combined magazine weights for these words decrease across year groups, reflecting both the decreasing overall magazine scores seen in Fig. 1 and the greater skew towards a small number of frequent types seen in Fig. 2. Importantly, this decreasing weight cannot be attributed to the nature of the words themselves since the mean magazine score remains almost constant across year groups. This implies that the lower overall weight is due to words' lower normed frequencies. Table 7 shows the mean and quartile normed frequencies for each year group. These show a clear decrease, such that older children use individual magazine-like types much less frequently than younger children. We can therefore conclude that the overall greater magazine weight of younger children's writing is not due to their using words that are more skewed towards this register, but rather to their repeatedly using a relatively small set of words.

The words in Table 6 also provide us with a means of understanding why magazine-like words are so much more prevalent in children's science writing than in other registers. We suggested above that this may be a product of magazines in COCA making prominent use of scientific vocabulary. Table 6 adds to this picture by showing in which category of magazine (as defined by the COCA corpus) each word is most frequently found. Across all 60 words, by far the most common category is *Health and Home* (41 words). These words are mostly used in COCA to discuss food (e.g. *flour; fruit; heat; microwave; rennet; stir*), health and nutrition (e.g. *anti-bacterial; calcium; cancer; diet; magnesium; sodium*), or gardening (e.g. *bulb; epiphyte; stem; texture*). A number of words are used prominently in two or more of these areas (e.g. *acid; bacteria; enzyme; potassium; protein; yeast; zinc*).

The second most frequent category is *Science and Technology* (10). More than half of these are predominantly used in the context of astronomy (*brightness; gravitational; infrared; meteor; meteorite; ultraviolet*). The others are either used in the fields of home electronics (*amp*) or the human body (*retina*) or else are general-purpose words found across a range of areas (*carbonate, microbe*).

The aim of this section has been to understand why magazine scores (a) are so much higher in children's science writing than in other disciplines and (b) decrease across year groups. In response to (a), we can now conclude that many magazines in COCA make prominent use of scientific vocabulary. This is partly due to popular science magazines, especially those focusing on astronomy. Surprisingly, a far greater influence is a combination of articles on food, health/nutrition, and gardening. This overlaps strongly with

the focus of the children's assignments, many of which focus on the topics of food, plants, and health.

In answer to (b), we have seen that it is the frequency with which magazine-like words are used, rather than the strength of their skew towards the magazine register, which influences the decrease across year groups. Although there is some change in the nature of the words used in that most non-scientific words were found at Year 6, such words are infrequent and their impact on overall scores relatively small. The key difference is rather that younger children use a small group of magazine-like words more frequently.

3.3. The use of academic words in Science and non-literary English

Fig. 1 demonstrated that academic vocabulary was the one measure that increased reliably across year groups for all text types; it also had the strongest MEM fit of any of the measures reported in Table 4 (conditional $R^2 = .88$). This echoes previous literature, which has demonstrated a consistent increase in academic vocabulary as children progress through school (Berman & Nir-Sagiv, 2007; Corson, 1985, 1989; Roessingh et al., 2016; Sun et al., 2010; White, 2015). Much less studied is the question of why this increase occurs. That is, what older children are doing in their writing that results in greater use of such vocabulary. Similarly neglected is the question of whether this increase is similar across different academic disciplines. To address these questions, this section will look more closely at the increase in academic vocabulary in the contrasting disciplines of Science and English (in the latter case, focusing on non-literary writing only).

As in the previous section, our starting point is to identify words which are characteristic of the register. This is again defined as words that have a register score more than one standard deviation above the mean for COCA. 17% of word types in the COCA list were classified as academic. Of these, 339 types were found in the children's Science texts and 596 in non-literary English texts. The full lists of types found for each year group are shown in Appendices B (Science) and C (English).

We saw in Section 3.2 that a small group of words accounted for most of the overall 'weight' of magazine vocabulary. As Figs. 3 and 4 show, this skew is much less pronounced in the current case.

Rather than restricting ourselves to a small number of prominent words as we did in the previous section, we therefore need to consider the full set of academic words to understand what drives the overall increase across year groups. Table 8 summarizes quantitative information for these words in Science writing. As with the magazine words, there was little difference between year groups in words' mean register scores (the small differences between year groups are all within a third of each group's standard deviation). However, there are large increases in the numbers of academic types and tokens. The range of different words (measured here using the Corrected Type-Token Ratio [CTTR], a measure of diversity which allows comparison across texts of different lengths⁶ (Carroll, 1964)) increased dramatically from Years 6–9, followed by a smaller increase from Years 9–11. Token frequencies increased steadily across year groups. We can therefore conclude that the overall increase in mean academic weight in Science writing is due to learners both employing a greater repertoire of academic words and repeating those words more frequently.

Table 9 provides a similar summary for academic words in non-literary English texts. Again, there is no increase in the mean academic score of types across year groups, so no clear difference in the nature of the words being used. The mean frequency of words also remains relatively constant (with the exception of Year 2, where type frequencies are very low and the token distribution strongly influenced by the frequencies of a small set of very frequent types). However, we see a large increase in diversity (CTTR) from year to year. In this case, therefore, it is an increasing repertoire of academic words that is driving the overall increase in academic vocabulary.

The next question is what these increasing repertoires of academic vocabulary consist of, and what functions they serve. To explore this, we retrieved all academic words that were attested only in Year 11 writing within the two disciplines. That is, those words which distinguished writing in the oldest year group from their younger counterparts. To avoid focusing on marginal cases, we looked only at words used in three texts or more. 21 types met this criterion in Year 11 Science writing and 72 in English writing. These types were examined in their original context using *CasualConc* (Imao, 2021) and categorized through an iterative, inductive process. The first author created an initial set of labels to describe the general function of each term and wrote provisional definitions for each. He then compared these labels and, where appropriate, combined them into broader categories or divided them into more fine-grained categories. Following this, the second author attempted to apply the codes and definitions to the full set of items and the two sets of codings were compared. Where first and second codings did not match, we discussed each item individually, exploring the reasons for differences and negotiating a final decision. This led to partial rewriting of some definitions and to an agreed coding for all items. The full set of words was then reviewed and categorized under these finalized categories. The outcome of this process was a set of twelve functional categories which effectively captured the main uses of the listed vocabulary in our corpus. These categories are summarized in Table 10 (for Science writing) and 11 (for non-literary English writing). Since the coding process showed that individual words often played multiple functions (sometimes simultaneously), we did not attempt to assign types to exclusive categories or to quantify the prevalence of each function.

While alternative taxonomies would no doubt be possible, something that emerges clearly from Tables 10 and 11, and that would remain regardless of the details of categorization, is that the academic words that distinguish Year 11 writing from that of younger students play very different roles in Science and English writing. The former are focused primarily on describing/discussing experimental procedures (examples 1 and 2) or naming/describing objects of scientific study (examples 3 and 4). The latter are used primarily to discuss set texts, focusing on their literary analysis (examples 6–8), discussion of their key events (examples 18 and 19) and ideas (examples 20 and 21), the meanings they express (examples 15–16), and their impact on the reader (examples 22 and 23).

⁶ CTTR is calculated as $\frac{\text{types}}{\sqrt{2 \times \text{tokens}}}$

Table 5
Prevalence of magazine-like words in Science writing.

	Total magazine-like words		% magazine words classified as <i>scientific</i>	
	Types	Tokens per 1 K words	Types	Tokens
Year 6	107	96.02	92	98.6
Year 9	81	52.18	95	99.4
Year 11	80	46.45	99	99.8

Academic words are also used to make comparisons between sources, ideas, viewpoints and characters (examples 9 and 10) and to give quotations from texts (examples 13 and 14). The one functional category that Science and English writing shared was that of *text organizer* – metadiscourse markers used to indicate the structure of the writer’s own text. However, these were represented by only a single word in each discipline (see examples 5 and 17).

The expanding academic vocabularies seen in Science and English are therefore responses to very different sets of communicative needs. This raises the question of how much academic vocabulary is common across year and discipline groups in general. To answer this question, Table 12 shows the percentage overlap in academic words used at each year group in each discipline. In general, Science and English writing have little in common: the highest cross-disciplinary overlap is found between Year 9 English writing and Year 11 Science writing, at 30%. There is also little overlap between primary (Year 6) and secondary (Years 9 and 11) school writing, regardless of discipline. Interestingly, the highest overlaps for the two primary-level groups are with each other, rather than with other groups within their discipline. By far the strongest overlaps are found between Years 9 and 11 within each discipline (40% for Science and 54% for English). Thus, while it is possible to talk of a relatively coherent set of academic vocabulary used in secondary English or in secondary Science, the two disciplines are relatively distinct from each other. There also seems to be a sharp division between primary and secondary use of academic words.

4. Discussion

Though vocabulary is widely acknowledged as a key area of language development for school-age children (e.g. Durham et al., 2007; Spencer et al., 2017; Townsend et al., 2012), previous research has studied this development through a restricted set of assessment types. Within studies of productive written vocabulary, we have particularly highlighted the lack of attention to register appropriateness. This is a key omission, both because vocabulary sophistication cannot be understood in isolation from register (words are only sophisticated in particular contexts) and because register appropriateness adds an aspect of vocabulary depth to measures of vocabulary use.

We recently proposed a quantitative measure of register appropriateness (Durrant & Brenchley, 2019). However, this was only applied to two target registers and our analysis did not take account of potentially important disciplinary differences in vocabulary use. Key questions were also left open regarding how quantitative differences in the measure should be interpreted.

The present study has shown that development can be observed in children’s use of vocabulary characteristic of all four registers studied. However, fiction and academic writing are much stronger developmental orientations than magazine and news, both in the sense that such vocabulary is more prevalent and in the sense that relationships between these registers and development are stronger.

Our register measure is ambiguous in the sense that it is unclear if quantitative differences across year groups reflect a qualitative change in the types of words that are used, a quantitative change in the extent to which certain types of words are used, or both. Analysis suggests that, at least with regard to magazine and academic vocabulary, the change is primarily quantitative. That is, the individual words used by older children are not, on average, *less magazine-like* or *more academic* in nature than those used by younger children. Rather, older children are less likely to frequently repeat magazine-like words and they use a larger number of academic words and repeat them more often.

This finding recollects our previous research (Durrant & Brenchley, 2019), which showed that the total repertoire of words (disregarding repetitions) used by young children included as much low-frequency vocabulary as that of older children. Indeed, it was the youngest children who used the lowest frequency nouns (e.g., *caldron*, *earworm*, *hideout*, *wisp*). What marked younger children’s writing as less sophisticated was that they repeated high-frequency words more often (interestingly, nouns remained an exception: younger children’s texts were marked by repetition of *low-frequency* words).

These findings have important implications for how we think about vocabulary development. A key distinction has been made in the literature between vocabulary diversity and vocabulary sophistication (e.g. Bulté, Housen, Pierrard, & Van Daele, 2008; Kyle, 2021; Leńko-Szymańska, 2020; Read, 2000). The former refers to the *size* of a learner’s vocabulary and is typically operationalized in

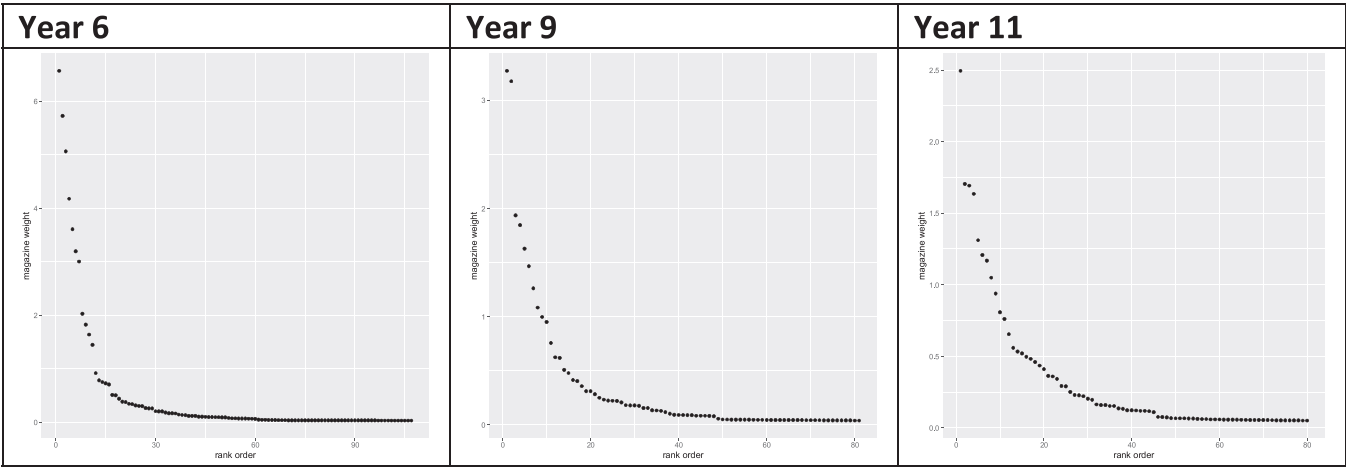


Fig. 2. Skew in magazine weights across words in Science writing.

Table 6
Top 20 words by magazine weight for each year group.

Year 6					Year 9					Year 11				
lemma	freq/ 1 K	mag. score	mag. weight	COCA category	lemma	freq/ 1 K	mag. score	mag. weight	COCA category	lemma	freq/ 1 K	mag. score	mag. weight	COCA category
yeast	11.13	0.59	6.57	H&H	cancer	6.65	0.49	3.27	H&H	enzyme	4.02	0.62	2.50	H&H
sugar	10.81	0.53	5.73	H&H	color	5.82	0.55	3.18	W/M	acid	3.15	0.54	1.71	H&H
microbe	10.13	0.50	5.06	Sci/Tech	acid	3.57	0.54	1.94	H&H	cheese	3.26	0.52	1.69	H&H
bacteria	8.19	0.51	4.18	H&H	magnesium	2.49	0.74	1.85	H&H	sodium	2.82	0.58	1.64	H&H
meteor	5.88	0.61	3.61	Sci/Tech	stem	3.24	0.50	1.63	H&H	carrot	2.60	0.50	1.31	H&H
bulb	6.56	0.49	3.19	H&H	aluminum	2.66	0.55	1.47	Sports	rennet	2.06	0.59	1.21	H&H
cup	5.56	0.54	3.00	H&H	powdered	2.33	0.54	1.26	H&H	uninsulated	1.74	0.67	1.17	Sports
flour	3.63	0.56	2.03	H&H	microwave	2.16	0.50	1.08	H&H	casein	1.30	0.81	1.05	Soc/Arts
amp	3.06	0.60	1.82	Sci/Tech	infrared	1.99	0.50	1.00	Sci/Tech	carbonate	1.74	0.54	0.94	Sci/Tech
mixture	2.81	0.58	1.64	H&H	diet	1.58	0.60	0.95	H&H	insulated	1.41	0.57	0.81	Sports
layer	3.06	0.47	1.45	H&H	heat	1.58	0.48	0.76	H&H	clot	1.63	0.47	0.76	H&H
brightness	1.75	0.52	0.92	Sci/Tech	ultraviolet	1.08	0.58	0.62	Sci/Tech	calcium	1.09	0.60	0.65	H&H
meteorite	1.56	0.50	0.78	Sci/Tech	squash	1.08	0.57	0.62	H&H	potassium	0.87	0.64	0.56	H&H
stir	1.38	0.55	0.75	H&H	zinc	0.91	0.56	0.51	H&H	color	0.98	0.55	0.53	W/M
tablespoon	1.38	0.53	0.73	H&H	pea	1.00	0.48	0.48	H&H	heat	1.09	0.48	0.52	H&H
tortoise	1.44	0.49	0.70	Sport	retina	0.83	0.50	0.41	Sci/Tech	texture	0.98	0.51	0.50	H&H
antibacterial	0.81	0.63	0.51	H&H	bulb	0.83	0.49	0.40	H&H	magnesium	0.65	0.74	0.48	H&H
croc	0.88	0.57	0.50	W/M	gravitational	0.75	0.48	0.36	Sci/Tech	protein	0.76	0.61	0.46	H&H
penguin	0.94	0.47	0.44	Soc/Arts	clot	0.66	0.47	0.31	H&H	thermostat	0.76	0.57	0.44	H&H
epiphyte	0.75	0.51	0.38	H&H	fruit	0.66	0.47	0.31	H&H	layer	0.87	0.47	0.41	H&H

H&H: Health & Home; W/M: Women/Men.

Table 7
Quantitative summary of top 20 magazine-weighted words in Science writing by year group.

	mag. weight		mag. score		freq/1 K					
	Total	Coverage	M	sd	Mean	0%	25%	50%	75%	100%
Year 6	43.97	84.92	.54	.05	4.09	.75	1.38	2.94	6.05	11.13
Year 9	22.40	79.95	.53	.06	2.09	.66	.89	1.58	2.53	6.65
Year 11	19.33	73.62	.58	.09	1.69	.65	.95	1.36	2.20	4.02

terms of the amount of repetition in their language use. The latter refers to the *nature* of the words that learners use. However, our findings suggest that it is misleading to think of individual word types as inherently sophisticated. Rather, sophistication is a property of a text as a whole and depends crucially on the variety of less frequent or more register appropriate words that a writer uses. The construct of sophistication therefore cannot be seen as strictly distinct from that of diversity.

We have also shown that distinctive courses of development are evident in different academic disciplines and text types. Specifically, the three disciplines of English, Science, and Humanities differed significantly in their levels of register-specific vocabulary and courses of development, as did literary and non-literary writing within the English subject area. Looking more closely at Science and non-literary English writing, we found that the repertoires of academic words differed sharply across the two disciplines and that these differences were driven by differences in the communicative demands to which such words were a response.

This suggests that children's academic vocabulary development at school should not be treated as a single process; different subject areas see very different types of development. Measures of vocabulary knowledge which do not account for such differences are therefore likely to be misleading. The absence of discipline as a variable in most previous school-age vocabulary research is therefore potentially problematic. It also emphasises how distinct parts of the curriculum pose distinctive communicative challenges and play different roles in the development of children's communicative competence. Written Science assignments, for example, give children the opportunity to use types of vocabulary and deal with types of communicative situation to which they are not exposed in English.

From a methodological perspective, we have emphasized the importance of following up broad-brush quantitative analyses with more fine-grained work that can recontextualize overall patterns. Leńko-Szymańska (2020) has argued for a more contextualized understanding of how vocabulary proficiency is reflected in vocabulary use, since the words that writers employ are a reflection not only of their proficiency, but of a myriad of complex contextual factors. Such contextual factors were seen clearly in the strong relationship between assignment topic and register measures in our mixed-effects models. It was also seen in the qualitative analysis of academic vocabulary, where the divergent communicative demands of different subject areas led to very different vocabulary profiles.

Extending Leńko-Szymańska's point, we can note that not only is the learner writing in our corpus strongly influenced by context, but so too is the reference corpus upon which our measures are based. We have seen how initially surprising patterns in the use of magazine-like vocabulary in children's Science writing cannot be properly understood without detailed study of the reference corpus. Understanding corpus-based measures therefore requires us to think about the contextualized nature of vocabulary use from the perspectives of both the learner corpus and the reference corpus.

5. Conclusions

This article has attempted to extend our understanding of development in children's written vocabulary and of how such development can be measured through the study of register appropriateness. We have shown that register measures demonstrate significant discipline- and genre-specific developmental patterns in but that these are strongest for measures of academic and fiction-like vocabulary. We would therefore propose that these two measures in particular represent useful additions to the tools available for assessing written language.

We have further shown that, in the case of two measures that were studied in detail, development was due to quantitative differences in the range and intensity of the use of register-appropriate words, rather than to changes in the types of words used. This suggests that the distinction between vocabulary diversity and vocabulary sophistication, which has been central to much corpus-based assessment of written language, may not be helpful for understanding children's L1 vocabulary development. From the beginning of their education, children use words that can, individually, be considered sophisticated. But as they mature, greater diversity, and decreased repetition, of word types creates a greater sense of sophistication across the course of a text.

One much-studied aspect of register-specific vocabulary is that of academic vocabulary. While our findings agree with previous research (Berman & Nir-Sagiv, 2007; Corson, 1985, 1989; Roessingh et al., 2016; Sun et al., 2010; White, 2015) in finding use of such vocabulary to increase over time, we have also shown that previous work has been simplistic in theorizing academic vocabulary as a single construct. Our results suggest that the development of academic vocabulary is discipline specific, with the academic vocabulary

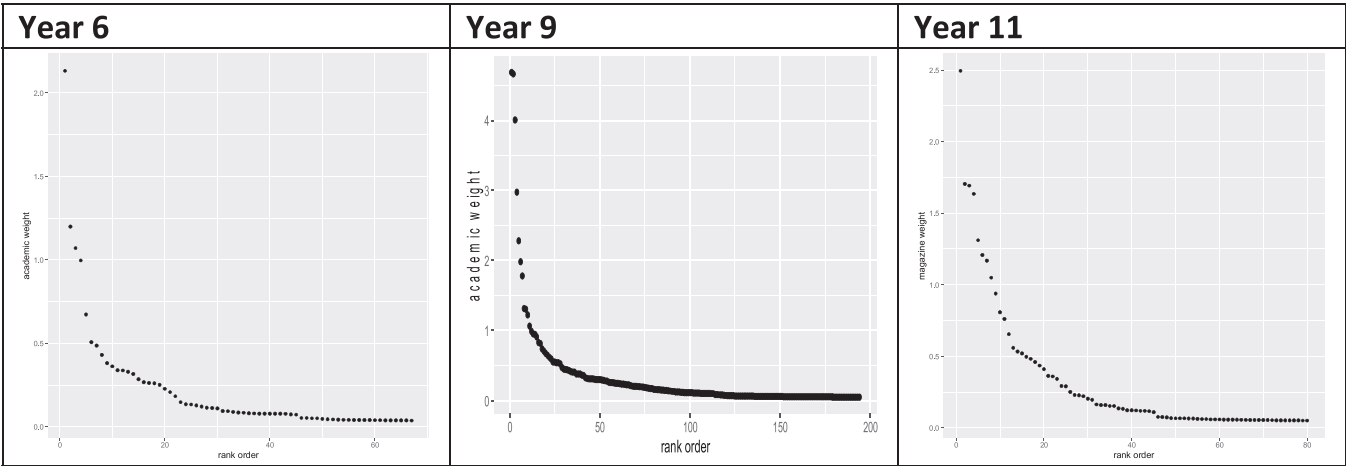


Fig. 3. Skew in academic weights across words in Science writing.

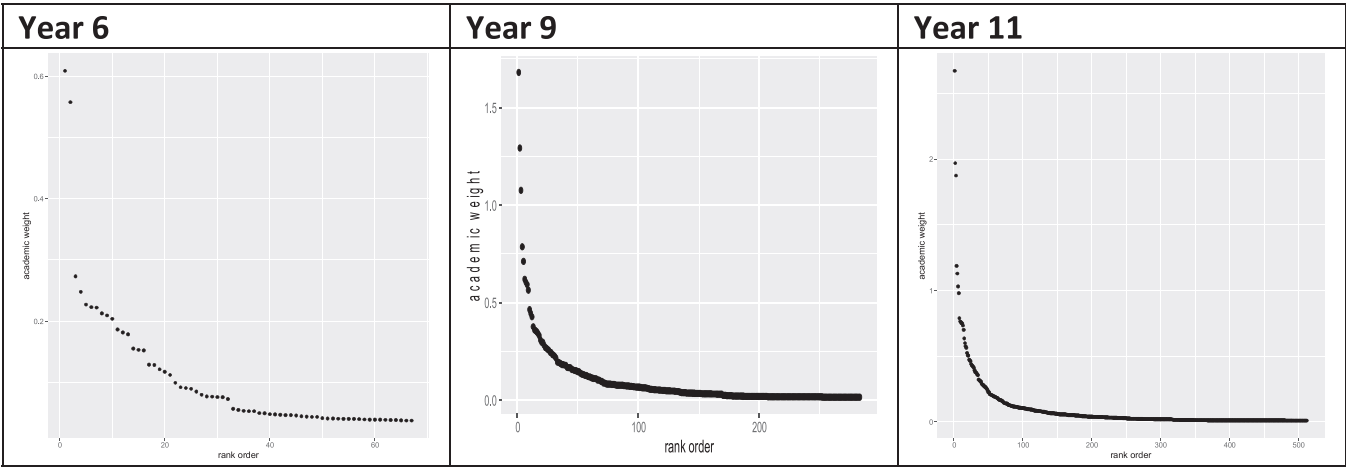


Fig. 4. Skew in academic weights across words in non-lit English writing.

Table 8
Quantitative summary of academic words in Science writing.

	acad. score		CTTR	freq/1 K					
	M	sd		M	0%	25%	50%	75%	100%
Year 6	.65	.07	.27	.33	.06	.06	.13	.38	2.59
Year 9	.68	.10	.90	.49	.09	.09	.18	.44	8.07
Year 11	.69	.10	1.00	.66	.11	.11	.22	.65	13.46

Table 9
Quantitative summary of academic words in non-literary English writing.

	acad. score		CTTR	freq/1 K					
	M	sd		M	0%	25%	50%	75%	100%
Year 2	.70	.08	.08	2.32	.21	.21	1.06	4.14	9.13
Year 6	.67	.08	.21	.16	.06	.06	.06	.19	.89
Year 9	.66	.08	.66	.15	.03	.03	.05	.13	2.86
Year 11	.66	.08	.98	.15	.02	.02	.03	.11	4.03

Table 10
Functions of academic words in Science writing.

Category	Definition	Examples
Describing/discussing an experimental procedure	Used to describe, discuss or structure an experimental procedure and/or instruments used in such a procedure	1. Our group also found it hard to keep our waterbath (brought to the correct temperature by mixing hot and cold water) at the correct temperature for the duration of the experiment. 2. Pour into the conical flask and add a drop of universal indicator
Naming or describing an object of study	Either refers by name to an object that scientists study or describes such an object or its behavior.	3. the water has travelled down the concentration gradient and into the carrot cells through the partially permeable cell membrane. 4. One similarity between the graphs is that they both show an increase in the biomass of zooplankton following the increase in biomass of the phytoplankton .
Text organizer	A metadiscourse marker, used to make the organization of arguments more explicit.	5. The advantages of comparing your results are to make sure you have no anomalies, to make sure it's reproductive and lastly to make see if there is any patterns.

of Science classes being quite distinct from that of English classes and functioning in response to very different communicative demands. These differences both point to the need for a more differentiated approach to assessing the development of academic vocabulary and highlight the importance of Science writing in the development of children's overall writing proficiency.

Finally, we have emphasized the importance of complementing broad-brush quantitative summaries of language use with more fine-grained study of the texts and reference corpora on which these summaries are based. Recent years have seen huge interest in measures of this sort, spurred on by the accessibility of easy-to-use computer applications. While this is to be welcomed, users of such measures need to be conscious of the complex effects that can underlie broad quantitative trends and of the careful analysis that is required to reach valid interpretations of them.

Table 11
Functions of academic words in non-literary English writing.

Category	Definition	Examples
Analysis	Describes or comments on the literary presentation of a text, often by citing a literary device or linguistic form/category that has been used	6. The use of the present participle allows the poem to flow and the words themselves create a sense of potential and almost frustration. 7. The poets differ in the way they structure their poems. 8. By use of the symbolism of the raven for death, Lady Macbeth immediately informs the audience of her intentions
Comparison	Compares sources, uses of language, images, viewpoints, or characters	9. But similarly to source A, it illustrates how the "picturesque" snow can completely cover areas. 10. Both sources reflect an image of the power of nature; contrasting between "destruction" and "beauty".
Non-textual situation or topic	Names or describes a topic, situation or idea not related to a set text	11. Many argue that snow causes disruption and increases the likelihood of accidents 12. We live in an era of technology, of progression, of development .
Quotation	Words quoted directly from a text	13. When he explains the age group away, he references that, ".a large proportion of these marriages are contracted between parties of very unequal ages" 14. Also we see guilt take on the theme of nature and unholiness when Macbeth says, "But let the frame of things disjoint , both the worlds suffer."
Stating meaning	Describes the meaning or function of a portion of text, either expressed directly by the writer or indirectly through the mouth of a character	15. Also, the image could connote that they have a vendetta. 16. The regularity in the structure of the poem could signify the stability of the father's relationship
Text organizer	A metadiscourse marker, used to make the structure of the text more explicit.	17. In summary , it is evident that Mr Birling is portrayed as a "hard-headed man of business", as he seems to have confidence in his opinions and views.
Textual episode	Describes or comments on a specific episode or situation that occurs in a text	18. At his daughter's engagement party he talks about how he "was Lord Mayor here two years ago" 19. The final paragraph shows how Alex reaches a sense of acceptance that "their mother's presence" was enormous.
Textual idea	Describes or comments on an idea or theme expressed in a text	20. While Sissy doesn't mention any principles of economic science she does show the core of her humanity and states a principle of society and socialist and communist economic models. 21. This is a presentation of the importance of social hierarchy in a 1984 post war British society.
Textual impact	Describes the effect of a text on an imagined reader or group of readers	22. The writer has now left the reader with the implicit thoughts of the child being kind, innocent and loved by all 23. Dickens uses many strategies and techniques to enhance the reader's sympathy for Louisa in this passage.

Table 12
Overlaps in academic words across text groups.

		Science			English non-literary	
		Year 6	Year 9	Year 11	Year 6	Year 9
Science	Year 6					
	Year 9	.25				
	Year 11	.23	.40			
English non-literary	Year 6	.27	.21	.21		
	Year 9	.19	.29	.30	.26	
	Year 11	.15	.27	.26	.17	.54

Acknowledgements

This research was supported by a grant from the Economic and Social Research Council (ES/M00967X/1). The Growth in Grammar corpus is openly available from UK Data Service ReShare at: <http://reshare.ukdataservice.ac.uk/853809>.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.asw.2021.100596](https://doi.org/10.1016/j.asw.2021.100596).

References

- Barton, K. (2020). *MuMIn: Multi-model inference, R package version 1.43.17*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes*, 43(2), 79–120.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.
- Bulté, B., Housen, A., Pierrard, M., & Van Daele, S. (2008). Investigating lexical proficiency development over time - The case of Dutch-speaking learners of French in Brussels. *French Language Studies*, 18, 277–298.
- Carroll, J. B. (1964). *Language and thought*. Prentice-Hall.
- Collins, B. A., & Toppelberg, C. O. (2021). The role of socioeconomic and sociocultural predictors of Spanish and English proficiencies of young Latino children of immigrants. *Journal of Child Language*, 48(1), 129–156.
- Corson, D. (1985). *The lexical bar*. Pergamon.
- Corson, D. (1989). Adolescent lexical differences in Australia and England by social group. *The Journal of Educational Research*, 82(3), 146–157. <https://doi.org/10.2307/40539585>
- Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415–443.
- Davies, M. (2008). <http://corpus.byu.edu/coca/> on 1 May 2018.
- Davies, M. (2018). *Word frequency data*. Retrieved November from (<https://www.wordfrequency.info>). November 2012.
- Department for Education. (2018). *Schools, pupils and their characteristics: January 2018*. Retrieved 20 April 2021 from (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719226/Schools_Pupils_and_their_Characteristics_2018_Main_Text.pdf).
- Durham, R. E., Farkas, G., Hammer, C. S., Tomblin, J. B., & Catts, H. W. (2007). Kindergarten oral language skill: A key variable in the intergenerational transmission of socioeconomic status. *Research in Social Stratification and Mobility*, 25, 294–305.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus annotation: Linguistic information for computer text corpora* (pp. 102–121). Longman.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. Longman.
- Durrant, P. (2014). Discipline- and level-specificity in university students' written vocabulary. *Applied Linguistics*, 35(3), 328–356.
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49–61.
- Durrant, P., & Brenchley, M. (2019). Development of vocabulary sophistication across genres in English children's writing. *Reading and Writing*, 32(8), 1927–1953.
- Durrant, P., Brenchley, M., & McCallum, L. (2021). *Understanding development and proficiency in writing: quantitative corpus linguistic approaches*. Cambridge: Cambridge University Press.
- Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10(1), 95–125.
- Hayes, J. R. (2012). Modelling and remodeling writing. *Written Communication*, 29(3), 369–388.
- Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–62.
- Hyland, K., & Tse, P. (2007). Is there an "Academic Vocabulary"? *TESOL Quarterly*, 41(2), 235–253.
- Imao, Y. (2021). *CasualConc (Version 2.1.6)*. In Osaka University. Retrieved from (<https://sites.google.com/site/casualconc/>). (Accessed 19 April 2021).
- Kyle, K. (2021). Lexis. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (pp. 332–344). Routledge.
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757–786.
- Leech, G. (2007). New resources, or just better old ones? The holy grail of representativeness. In M. Hundt, & N. Nesselhauf (Eds.), *Corpus linguistics and the web* (pp. 134–149). Rodopi.
- Leńko-Szymańska, A. (2020). *Defining and assessing lexical proficiency*. Routledge.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496.
- Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, 8, 1442.
- Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing*, 22, 545–565.
- R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Roessingh, H., Douglas, S., & Wojtalewicz, B. (2016). Lexical standards for expository writing at grade 3: The transition from early literacy to academic literacy. *Language and Literacy*, 18(3), 123–144.
- Roessingh, H., Elgie, S., & Kover, P. (2015). Using lexical profiling tools to investigate children's written vocabulary in grade 3: An exploratory study. *Language Assessment Quarterly*, 12(1), 67–86.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951.
- Shaw, S. D., & Weir, C. (2007). *Examining writing: Research & practice in assessing second language writing*. UCLES/Cambridge University Press.
- Spencer, S., Clegg, J., & Stackhouse, J. (2012). Language and disadvantage: A comparison of the language abilities of adolescents from two different socioeconomic areas. *International Journal of Language & Communication Disorders*, 47(3), 274–284.
- Spencer, S., Clegg, J., Stackhouse, J., & Rush, R. (2017). Contribution of spoken language and socio-economic background to adolescents' educational achievement at age 16 years. *International Journal of Language & Communication Disorders*, 52(2), 184–196.
- Sun, Y., Zhang, J., & Scardamalia, M. (2010). Knowledge building and vocabulary growth over two years, Grades 3 and 4. *Instructional Science*, 38, 147–171.
- Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics*. Pearson.

- Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal*, 112(3), 497–518.
- White, R. (2015). The expanding literate lexicon in L1 secondary school academic writing: A mixed methods study. *New Zealand Studies in Applied Linguistics*, 21, 5–20.
- Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

Philip Durrant is Associate Professor in Language Education at the University of Exeter. He has been a language teacher and researcher for over twenty years, working at schools and universities in both the UK and Turkey. He has published widely on corpus linguistics, vocabulary learning and academic writing.

Ayça Durrant is co-ordinator of the English for Academic Purposes Foundation Programme, at INTO, University of Exeter. She holds an M.A. in Language Testing from Lancaster University and has taught English from primary school to university levels in the UK and Turkey.