Journal of English for Academic Purposes 43 (2020) 100804

Contents lists available at ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Investigating the effects of reducing linguistic complexity on EAL student comprehension in first-year undergraduate assessments

Daniel Riccardi [a, *], Jennifer Lightfoot [a], Mark Lam [b], Katherine Lyon [c], Nathan D. Roberson [d], Simon Lolliot [b]

[a] University of British Columbia, UBC Vantage College, Canada
[b] Department of Psychology, University of British Columbia, Canada
[c] Department of Sociology, University of British Columbia, Canada
[d] Department of Education and Counselling Psychology and Special Education, University of British Columbia, Canada

ARTICLE INFO

Article history:
Received 11 January 2019
Received in revised form 2 October 2019
Accepted 7 October 2019
Available online 9 October 2019

Keywords:
Assessment
EAP
Undergraduate
First-year
Complexity
SFL
Multiple choice questions

ABSTRACT

Academic writing across disciplines is often linguistically complex, characterized by abstract ideas densely packed into nominal groups (Biber & Gray, 2010; Halliday & Martin, 1993; McCabe & Gallagher, 2008), along with infrequent lexis and content requiring specific cultural knowledge. This linguistic complexity presents a significant comprehension challenge, contributing to an increase in the performance gap between English as an additional language (EAL) students and their non-EAL peers (Abedi & Gándara, 2006). This study presents the outcome of a collaborative project between Psychology, Sociology, and EAP instructors teaching within a pathway program at a Canadian university combining first-year university courses with language-linked EAP courses. One key outcome of this collaboration has been greater awareness of the comprehension challenges that assessments pose for students, particularly in the case of multiple choice question (MCQ) exams. To investigate the effects of linguistic complexity, the research team analyzed whether unpacking MCQs by reducing the linguistic complexity in test questions improves comprehension for EAL students. Our findings indicate that EAL students are more likely to score higher on unpacked assessment questions, highlighting the importance of reducing the complexity of language in assessments to provide linguistic space for novice students to demonstrate their knowledge of disciplinary content.

1. Introduction

The transition into first-year university can often be challenging, with students adjusting to greater academic demands, managing their own time, forming social support networks, and generally taking more responsibility for their learning. This transition may be particularly challenging for first-year international students, especially for English as an additional language (EAL) students (Andrade, 2006; Hyland, 2004). A main part of this challenge is due to the intense comprehension

* Corresponding author. The University of British Columbia, Vantage College, Orchard Commons - Room 3021, 6363 Agronomy Road, Vancouver, BC, V6T 1Z4, Canada.
E-mail address: daniel.riccardi@ubc.ca (D. Riccardi).

https://doi.org/10.1016/j.jeap.2019.100804
1475-1585/

demands of academic language. Research has shown that academic writing across disciplines is generally considered linguistically complex (Biber & Gray, 2010; McCabe & Gallagher, 2008; Staples, Egbert, Biber, & Gray, 2016). Previous studies in this area have mainly analyzed clausal complexity, with a focus on the measurement of T-units: the combination of dependent subordinate clauses connected to an independent clause (Ellis & Yuan, 2004; Vyatkina, 2012). However, recent research has focused on phrasal complexity, based on evidence that academic writing is more characterized by nominal group density rather than clausal complexity (Ansarifar, Shahriari, & Pishghadam, 2018), placing high importance on nominalization (Fang, Schleppegrell, & Cox, 2006; Halliday & Martin, 1993), and characterized by abstract ideas densely packed into nominal groups, making meaning less explicit for the reader (Biber & Gray, 2010; McCabe & Gallagher, 2008; Staples et al., 2016).

This linguistic complexity in academic discourse not only presents a significant comprehension challenge, but also contributes to an increase in the performance gap between EAL students and their non-EAL peers on assessments (Abedi & Gándara, 2006; Abedi & Lord, 2001). As student success in academia requires an ability to understand and make meaning within specific disciplinary contextual practices, instructors are faced with the challenge of making concepts written in a highly abstract manner intelligible to novice students (Moore, 2007). Most of these studies on clausal and phrasal complexity have focused on student writing, specifically the development of complexity in their writing, as opposed to comprehension of this linguistic complexity.

In response to this challenge, a growing number of higher education institutions have attempted to support EAL students through the creation of a wide variety of language supported educational programs. This support ranges from a “study skills” approach, generally centred around short workshops and courses that are often delivered by university support centres, to a more “academic socialization” approach guided by an explicit instructional focus on teaching learners about the key features of the relevant genres and text types they will encounter (Hyland, 2006), often in the form of full semester university courses. This socialization may take the form of EAP courses that seek “to prepare students for a wide variety of target situations or English for specific purpose offerings,” which “tend to be designed in consultation with discipline specialists and are informed by a genre analysis of relevant assessment tasks” (Storch, Morton, & Thompson, 2016, p. 479).

This study was conducted within an innovative credit-bearing pathway program at a university in Western Canada that combines first-year university courses of various disciplines relevant to students' programs with language-linked EAP courses. Within this program, EAP instructors work closely with instructors of various disciplines to deliver custom-designed language courses that provide ongoing language support relevant to the specific courses and disciplines students study. The EAP courses are guided by a Systemic Functional Linguistics (SFL) framework. Motivated by the influence and pedagogical applicability of SFL and Halliday's functional grammar in educational contexts, this focus on register is intended to help students with the challenges posed by disciplinary discourse (Moore, 2007). The application of SFL theory along with the close collaboration between instructors in the program aim to help students identify, comprehend and adopt disciplinary practices that allow students to acquire and demonstrate their knowledge of course content in the various disciplines they study (Ferreira & Zappa-Hollman, 2019).

One key outcome of these collaborations has been a greater awareness of the comprehension challenges that assessments pose for students, particularly in the case of multiple choice question (MCQ) exams. Psychology and Sociology instructors teaching within the program have reported that the EAL students from our transitions program require additional time to complete MCQ exams, ask more clarifying questions, and score lower on such exams compared to students admitted directly into first-year UBC programs. These observations of differences in student performance are supported by previous research suggesting that language use in assessments that is above the proficiency level of learners can pose a cognitive burden and lead to lower student scores, specifically as a result of the linguistic complexity of test items (Abedi, Hofstetter, & Lord, 2004; Abedi & Lord, 2001; Parkes & Zimmaro, 2016). At the same time, this research suggests that reducing the linguistic complexity of test items improves student performance, helping to decrease the performance gap between EAL and non-EAL students. In their evaluation of linguistic complexity in MCQs for elementary school students assessing their mathematics knowledge, Abedi and Lord (2001) found that student performance improved when the complexity of the language used in the MCQ items was reduced, especially in the case of EALs and lower performing students.

In order to reduce this cognitive burden in MCQ exams, Parkes and Zimmaro (2016) call for language use “appropriate to your students' academic level and position relative to the profession” (p. 24) as it would be an unfair assessment of content knowledge if the reading level is beyond the students. Despite the call for increased language based accommodations on assessments for EAL students (Rivera, Stansfield, Scialdone, & Sharkey, 2000), Abedi et al. (2004) found that only a small percentage of test accommodations included reductions in linguistic complexity. Expanding on previous research, the present study reports the outcome of a collaborative project between the disciplinary instructors teaching first-year Psychology and Sociology courses as well as EAP instructors teaching within the university program. Given the prominence of the MCQ in higher education assessment (Haladyna, Downing, & Rodriguez, 2002) and the previously discussed research connecting a performance gap in assessments to complex language use, the research team analyzed whether unpacking MCQs by reducing the linguistic complexity in test questions improves student performance on test items.

Matruglio, Maton & Martin's (2013) work on semantic gravity depicts how instructors tend to already be very practiced at helping students unpack disciplinary content. Instructors often help students unpack key course concepts by providing more concrete examples and explanations of abstract, discipline-specific concepts through the use of more colloquial language in the classroom, which tends to be more easily applicable and understood. In comparison to the academic written discourse students are expected to engage in, language use revolving around key concepts in the classroom is typically identified and unpacked to aid understanding and learning. Given this natural attention to using more unpacked forms of language in the

classroom, along with the relative lack of emphasis on reducing the linguistic complexity of written assessments (Abedi et al., 2004), the essence of the present study is to rewrite MCQs in order to make the meaning of the question more accessible to all students so they can better understand the question and thus better demonstrate their content knowledge.

This intervention must be understood within the context of our program, particularly with the academic language support in mind. The foremost goal of this study is to determine the usefulness of this intervention on student test performance; however, the broader goal of our program is to produce learners who are capable of not only comprehending complex academic discourse but who are also able to produce it themselves. In the next section, we will review the three main linguistic interventions used in this study: reducing nominal group complexity, vocabulary substitutions and making cultural references more explicit. We will also discuss how this broader goal of our program is addressed in the academic language courses.

1.1. Nominal group complexity

Clear linguistic differences exist between spoken and written language (Biber, Gray, & Poonpon, 2011; Halliday, 1987). Halliday states that spoken language is generally characterized by higher clausal complexity whereas writing relies on higher lexical density to communicate messages with more frequent use of longer nominal groups. Whereas spoken language tends to be focused around the verbal group, resulting in greater clausal complexity, Halliday (2007) explains that the nominal group, which "construes reality as entities (objects, including instructional and abstract objects and their quantities, qualities and types)" is the central focus of written language (p. 109). This focus, particularly in the case of academic writing, results in a more abstract representation of ideas through more sophisticated use of nominal groups (McCabe & Gallagher, 2008).

When analyzing nominal groups using functional grammar from an SFL perspective, Halliday and Matthiessen (2014) direct attention to how a head noun (thing) can be pre- and post-modified. The pre-modification structure of a nominal group can consist of Deictics (determiners), Numeratives (numerals), Epithets (most often adjectives), and Classifiers (most often nouns or adjectives) while Qualifiers, such as prepositional phrases, embedded clauses and embedded non-finite clauses, can serve as post-modification of the head noun, additionally packing more information into the nominal group. Fig. 1 shows how each of these elements can be distinguished within a densely packed nominal group:

As readers transition through various stages from young learners in primary, elementary and high school, to undergraduate university students and experts in a specific discipline, they encounter increasingly dense nominal groups which place progressively greater demands on comprehension (Thompson, 2014). This complexity in academic texts as a result of high lexical density and nominalization poses significant reading comprehension challenges for EAL students as meaning becomes difficult to unpack (Ventola, 1996).

1.2. Vocabulary and cultural comparisons

Along with the difficulty of unpacking dense nominal groups, international EAL students face challenges deciphering infrequent lexis and content requiring specific cultural knowledge. The greater amount of unfamiliar vocabulary in academic texts increases reading comprehension difficulties, which may result in an inability to identify and understand the main ideas and significant details of a text (Hu & Nation, 2000). Hu and Nation suggest that learners must be able to understand as much as 98% of a written text in order to adequately comprehend it. In their review of the vocabulary literature, Hacking and Tschirner (2017) explain that it is common for learners to progress from higher to lower frequency words as their reading comprehension improves, citing "robust correlations between vocabulary knowledge and reading proficiency" (p. 513), concluding that "authentic literary texts are thus mainly beyond the reach of all but the most advanced students" (p. 515). In response to this challenge, Hacking and Tschirner note recent efforts that aim to select vocabulary appropriate to student proficiency levels in order to increase student comprehension and advance reading proficiency. Given that the minimum proficiency level of students in our program is 70 on the Test of English as a Foreign Language (TOEFL) or 5.5 overall on the

the	many	widespread	cultural	norms	that we use	to evaluate ourselves
Deictic	**Numerative**	**Epithet**	**Classifier**	**Thing**	**Qualifier** embedded clause	**Qualifier** non-finite embedded clause

Fig. 1. Nominal Group Structure Example.

International English Language Testing System (IELTS) entry requirement, these EAL students are certainly still in the earlier stages of developing their academic vocabulary.[1]

In addition, instructors often use cultural references as "hooks" to exemplify and situate course concepts in hopes of intriguing students; however, for international students, these cultural references can be an additional barrier to decode and interpret (Lee, 1997). These references often exist in the form of infrequent vocabulary that is context-sensitive, compounding the challenges in not only understanding lexical items, but also in the comprehension of the surrounding cultural context, which can greatly hinder a learner's ability to make meaning (Hsu & Yang, 2013). Duff and Zappa-Hollman (2012) emphasize the significant cultural and linguistic background knowledge required to understand popular culture references in the language classroom and the importance for instructors to be aware of the comprehension challenges these pose for EAL learners and newcomers. Furthermore, international students often do not recognize the cause of their confusion as a result of a lack of contextual knowledge, and instead often think it is their language skills that are causing the problem (Andrade, 2006), which may prevent students from asking for clarification about cultural references.

1.3. Present study

To investigate the effects of linguistic complexity on MCQ test performance for EAL students, five complex MCQs were identified and unpacked in midterm and final tests over two university terms and three first year courses: two Psychology courses and one Sociology course. A combination of three criteria were identified as contributing to the complexity of MCQs: nominal group complexity, vocabulary frequency as determined through the Corpus of Contemporary American English (COCA), and cultural knowledge specific to North America. This cultural knowledge was identified by two EAP instructors in the program based on 7–10 years of experience teaching EAL learners in higher education institutions. Once the complex MCQs were identified and unpacked, students in Psychology and Sociology courses received both complex (original) and unpacked (revised) versions of MCQs on each of their exams throughout two terms. Our goal was to determine whether such unpacking impacts comprehension, and therefore performance on unpacked test items. Thus, this study attempts to address the following research question:

To what extent does unpacking multiple choice questions support EAL students' ability to comprehend and demonstrate their knowledge?

2. Methods

2.1. Context & participants

In order to answer this research question, we collected data from a group of first-year EAL international students completing an Arts or Management degree who attended a specialized pathway program at the University of British Columbia in Western Canada. Students attending this program are academically strong and meet the competitive scholastic requirements for entry to the university. They do not, however, meet the linguistic requirements for admission typically scoring a minimum of 70 on TOEFL or achieving a minimum overall score of 5.5 on IELTS. These students thus complete their first full year at this university in either Arts, Engineering, Management, or Science, with additional, embedded support in academic English. As such, students take content courses (e.g., Psychology, Sociology) that are complemented by small-group classes, workshops and individual tutorials in academic English. The academic English support that the students receive is informed by an SFL approach which encourages students to analyze how different disciplines use language to construct and communicate knowledge. Upon successful completion of this pathway program, these students are then admitted into the general university population with second-year standing. The program welcomes students from over 50 countries with students from China comprising the largest enrollment.

The EAP courses support students in a wide variety of ways, with a focus on developing student awareness of the linguistic features of the particular disciplines they are studying as well as their ability to produce knowledge (i.e., through speech and writing) within these disciplines at a first-year, undergraduate student level. A wide variety of carefully sequenced topics are included in the curriculum in order to achieve this goal, such as academic reading, lecture listening and note-taking strates, along with the development of written skills, such as improving paragraph structure, flow, paraphrasing, summarizing, appropriate citation practice, and making claims appropriately through the communication of attitude and evaluation. There is also an important emphasis on planning for, drafting and proofreading written texts, with instruction and associated assessments carefully sequenced within the curriculum. Students are also provided significant opportunities to develop public speaking as well as small group and whole class discussion skills in an academic context.

Given the prominence of abstract ideas expressed through densely packed nominal groups in academic written discourse (McCabe & Gallagher, 2008), the nominal group itself and the process of nominalization is given particular emphasis in the language courses. As Matruglio, Maton, and Martin (2013) point out, most experienced instructors are already very skilful at unpacking key concepts to students, but a key requirement of university writing is to also then re-pack and condense

[1] More specifically, students may not score below 16 on any section of the TOEFL or, for the IELTS entry requirement, students must achieve a minimum band score of 5.0 in the speaking and listening components and 5.5 in reading and writing.

information; therefore, the curriculum aims to have students' language match the complexity of academic written discourse. Our language classes direct students' attention to this early in the curriculum, with extensive support on learning to identify nominal groups and observing how packing and unpacking occurs as language is translated across written and spoken forms, generalist and specialist audiences as well as other registers, such as between disciplines. With formative feedback provided on written drafts, and explicit assessment of nominal groups in associated final written assignments as well as midcourse and final tests, students' abilities to both comprehend and produce dense academic language are expected to improve following two university terms of instruction and practice.

Thus, the ultimate goal of our curriculum is student production of complex academic written language, demonstrated through course assessments. At the same time, however, as our students are completing first-year undergraduate courses with language proficiency levels below that of their direct entry peers in the same courses, we endeavour to provide them with the greatest possible opportunity to demonstrate their content knowledge on assessments. Assessments should not be designed to unintentionally test language proficiency (Abedi et al., 2004) but rather should prioritize the testing of content knowledge.

2.2. Unpacking test items

Throughout two university terms, Psychology and Sociology instructors provided the EAP instructors with the original MCQ test items prior to each test date. The EAP instructors then analyzed the MCQ test item stems, the part of the MCQ test item that identifies the question or problem, for complexity based on the three criteria described above: nominal group density, the inclusion of less frequent vocabulary, and ideas that appeared to require some degree of cultural knowledge specific to North America. The EAP instructors then identified and selected five complex MCQ stems per test based on these three criteria, and unpacked them to reduce the complexity in three ways.

First, longer, more complex nominal groups were expanded following Halliday and Matthiessen's (2014) framework. As explained above, academic written language is often more difficult to understand than spoken language as "typically, written language becomes complex by being lexically dense: it packs a large number of lexical items into each clause; whereas spoken language becomes complex by being grammatically intricate" through the use of clause complexes (Halliday & Matthiessen, 2014, p. 726). Also, Thompson (2014) argues that lexical density tends to increase with greater disciplinary specialization. Following Halliday, Matthiessen and Thompson, nominal groups were unpacked, rendering the sentences in the MCQ test stems more explicit by reducing phrasal complexity and increasing clausal complexity. In their study on grade 8 student test performance, Abedi and Lord (2001) reduced the complexity of MCQ math test items in a variety of ways, including avoiding passive voice, conditional clauses, relative clauses, and reducing the length of nominal groups. However, in the present study, nominal group density was specifically targeted in the unpacking of test items in order to determine the effect of this type of complexity in combination with lower frequency vocabulary and the inclusion of cultural references.

Abedi and Lord (2001) also highlighted low frequency vocabulary as an impediment to student comprehension of test items. Thus, in addition to reducing nominal group complexity, less commonly used academic vocabulary was also identified and substituted for a more frequently used synonym where possible through frequency checks using COCA. Finally, any questions which included references requiring specific North American cultural background knowledge were also explicitly explained. It is worth noting, however, that unpacking nominal groups and replacing infrequent lexis were by far the main strategies used to reduce the linguistic complexity of MCQ question stems, with explicit explanations of cultural references comprising a small minority of items.

In most questions, lexical density was reduced and clausal complexity was increased as nominal groups were unpacked. Halliday and Matthiessen (2014) calculate lexical density by dividing the number of lexical items by the number of clauses. In the example below, following Halliday and Matthiessen's approach, the lexical density of the complex version of the MCQ stem, which contains only one clause, is 14. On the other hand, the lexical density of the unpacked version, containing four clauses, is 5.75, and thus contains greater clausal complexity.

Example 1. **Complex MCQ:** The class of psychological disorders characterized by people being deprived of contact with portions of their consciousness that results in a disruption in their sense of identity is:

Unpacked MCQ: There are different classes of psychological disorders. In one class of psychological disorders, people lose contact with portions of their memory. This loss can disrupt their sense of identity. What is the name of this class of psychological disorders?

For many questions, we chose to both unpack a nominal group and replace some of the less frequent lexis with more frequent synonyms where possible through conducting searches on the COCA online interface, as in the example above which substitutes *deprived of contact* with *lose contact*. *Consciousness* was also replaced with *memory*, as *memory* occurs more frequently than *consciousness*, keeping in mind that the context of the course material allowed us to refer to this concept using either of these terms.

Reducing nominal group complexity and substituting vocabulary were often completed simultaneously, as unpacking typically involves word substitution. This approach to unpacking nominal groups is in line with SFL, "which postulates lexicogrammar as a sole complex object of linguistic inquiry instead of the two separate layers of grammar and lexicon" (Vyatkina, 2012, p. 578). This is similar to the changes made by Abedi and Lord (2001), who often made multiple alterations to single test items according to their complexity. In the following second example, unpacking the nominal group *raise the*

standing required greater expansion and the use of more frequently used synonyms to help make meaning more explicit for student comprehension.

Example 2. **Complex**: Du Bois's work to raise the standing of Afro-North American people closely follows the approach to social problems that would be taken by:

Unpacked: Du Bois's research tried to improve the social, political and economic status (position) of Afro-North American people. Other sociologists have also taken this approach to try to solve other social problems. Whose approach to social problems is similar to Du Bois's research?

In many cases, we noticed it was optimal to both unpack nominal groups while also substituting more frequently used synonyms to make meaning as explicit as possible. In the third example below, lexically dense nominal groups such as *significant factors that determines the number of children a woman bears* becomes *has the strongest effect on the number of children women have*. As can be seen in this example, substituting the infrequent verb *to bear children* is replaced with the more frequently used synonym *to have children* solely to aid student comprehension rather than to unpack the nominal group.

Example 3. **Complex**: From a sociological perspective, one of the most significant factors that determines the number of children a woman bears is:

Unpacked: Certain factors in society can cause women to have more or fewer children. The average number of children women have is different in different societies. From a sociological perspective, which of the following factors has the strongest effect on the number of children women have?

Finally, in a minority of test items, cultural references, specifically ones pertaining to Canadian culture, were explained explicitly. This was typically combined with unpacking strategies, such as in the fourth example below.

Example 4. **Complex**: Charon argues that the slap of a beaver's tail is:

Unpacked: Beavers are animals that have large, round tails and sometimes slap (hit) them on water. Charon argues that when a beaver slaps its tail, it is:

2.3. Data Collection Procedure

The research team conducted the study across six separate classes of introductory (100-level) Sociology and Psychology courses, three classes in the fall 2017 term and then an additional three classes in the spring 2018 term. In each term, students were assessed with three midterms and a final exam in each course. For each exam, five questions were selected and revised by the EAP instructors through the steps discussed above. Each exam included both the revised and original versions of these five questions, and all students completed all question versions through two-stage randomized exams.

As Fig. 2 shows, each exam had two versions with the unpacked and complex questions in opposite question order. For the first 5 minutes of each exam, students completed exam part 1 only (the supplement), consisting of a set of five questions on a separate cover page. Every question in this supplement was part of the research data. For every exam, there were two different versions of this supplement as explained above. Half of the class received exam supplement version A (consisting of Q1 unpacked, Q2 complex, Q3 unpacked, Q4 complex, Q5 unpacked), and half of the class received exam supplement version B (consisting of Q1 complex, Q2 unpacked, Q3 complex, Q4 unpacked, Q5 complex). Students could not view or commence the remainder of the exam until they completed this supplement and it had been collected by a member of the teaching team. This procedure ensured that students would not have access to the complex and unpacked versions of the same question at the same time. The two-stage randomized exam procedure was also implemented to account for the effect of receiving question types in a certain order. For any given MCQ, half of the class completed the complex version first and the other half of the class completed the unpacked version first.

All exam questions counted toward students' exam score for the course.[2] However, only students who opted into the study had their aggregated, anonymized MCQ responses analyzed. Each instructor in the study used the same Scantron template to administer MCQ exams to students. Scantron is a device that efficiently scans and calculates student responses to MCQs, which they complete on paper forms during the test. For the purposes of course grading, instructors recorded responses to each MCQ by student number in Microsoft Excel. These Excel files were then anonymized and aggregated by the project research assistant, who analyzed responses only for the unpacked and complex versions of the five questions initially selected by the EAP instructors (i.e., not the additional questions in the exam which did not have unpacked versions and thus were not included in the research study as described above).

Given that the Sociology and Psychology instructors were researching their own students, important ethical questions arose regarding power relations and possibilities of coerced consent. All communications about this study were conducted by a research assistant who introduced the study to students by email, visited each class to offer an overview of the project, and administered and managed consent forms. Instructors were never in the classroom when the research project was being

[2] Note that each instructor determined the weight they awarded to the first five MCQs in the exam part 1 supplement. For example, the sociology instructor awarded a grade of 0.5 for each of these five supplemental MCQs answered correctly, and a grade of 1 for each MCQ answered correctly in the remainder of the exam. This had a small influence on the overall course grade, as students could potentially receive a maximum of 2.5 bonus marks for completing the supplement, but has minimal implications for the findings of the study.

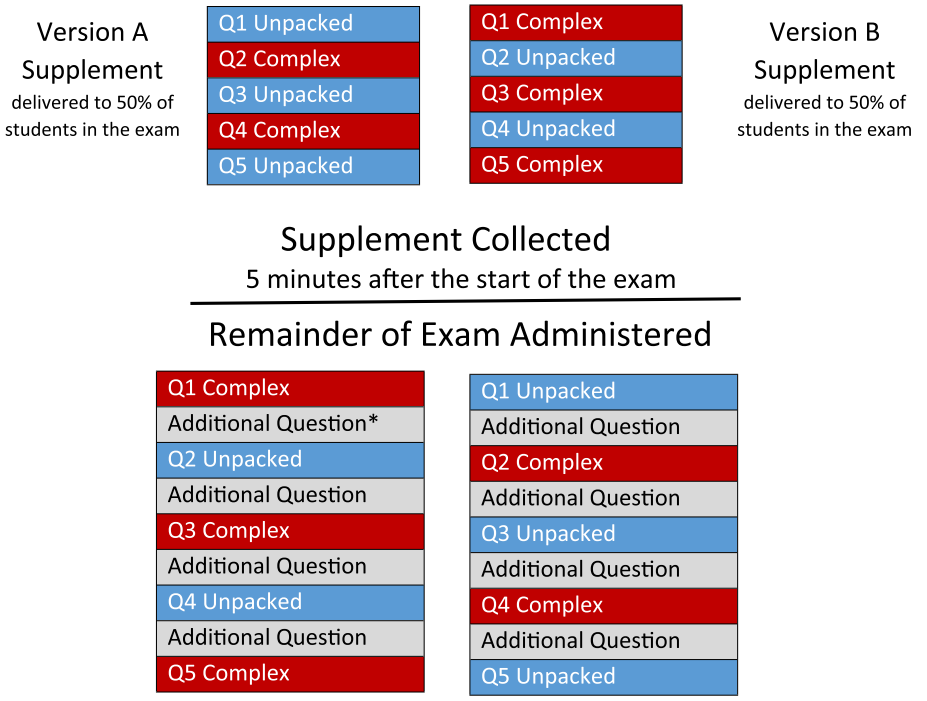

Fig. 2. Data Collection Procedure.
* Students also completed questions that did not have unpacked versions included in the exam, meaning that these test questions were not altered from their original form in any way. We refer to these questions as "additional questions" in Fig. 2. As these questions were not unpacked by EAP instructors, these questions were not analyzed as data. Only questions with two versions, complex and unpacked, were compared and analyzed in this study.

discussed with students. Instructors were also not aware which students gave consent for participation in the project. To respect students' limited time, the study was also designed to ensure that participation did not require additional effort beyond what was expected of students for their course assessments. The study received approval from the Behavioural Research Ethics Board at the associated university and was designated as a minimal risk study.

2.4. Analysis

To estimate the effect of unpacking on students' ability to demonstrate their knowledge on MCQs, we fit a multi-level logistic regression model controlling for student across time. Given that our primary research question is about assessment question wording, we used test items as the unit of analysis, not students. Each test item, irrespective of being unpacked or not, has a unique level of difficulty based upon content of the question and student interaction with the question. To use students as the unit of analysis would make the assumption that all items are equally difficult across time and students; as such, it is more appropriate to focus on the item level.

Table 1 details the number of items and students that were included in the study. For our analysis we have a total of eight test periods possible across one year (two terms); however, not all students will have data for all test periods since the number of testing periods is defined by individual instructors. As such, this provides further justification for the inclusion of time period as a necessary nesting variable. For each $item_i$ of $student_j$ at $time_k$ we estimated the odds ratio of a student answering the item correctly if the item was unpacked or not (Unpacked = 1, Complex = 0). For our analysis, the odds ratio for "unpacking" is the ratio of answering a question correctly if the question is unpacked. Odds ratio values are on a scale of 0 to positive infinity. An odds ratio of 1 corresponds to no effect, meaning a student would be equally likely to answer either an unpacked or complex question correctly. Odds ratio greater than 1 corresponds to a positive effect, and values between 0 and 1 correspond to a negative effect. Main effects were tested using a z-test ($\alpha \leq .05$). When values are greater than 1, odds ratio are defined as the number of times more likely an event will occur compared to the control. For example, an odds ratio of 2 would suggest an event is 2 times more likely to occur. A unique property of positive odds ratio values between 1 and 2 is that the decimal value simply represents the increased likelihood in percentage point values (e.g., 1.50 would be 50% more likely), which is equivalent to expressing the relationship in odds (e.g., 1.50 would mean someone is 1.50 times more likely to answer correctly).

There were a total of three instructors included in our study; thus, we included two indicator variables of instructor (instructor A and instructor B, coded 1 = if student had that instructor and 0 otherwise). Instructor effects are considered "fixed-effects" in our model, meaning the classroom effects for $item_{ijk}$ is held constant. Random effects for student (level 2)

Table 1
Number of test items & students included in testing periods.

Period	# items	# students
0	1,300	130
1	2,742	250
2	2,430	244
3	1,290	129
4	2,440	244
5	2,320	232
6	350	35
7	2,370	237

Table 2
Results of the effect of unpacking reported in logit odds.

Correct	Logit		Odds Ratio	
	b/(se)		or/(se)	
Unpacked	0.08	**	1.08	**
	(0.04)		(.04)	
Instructor A	0.48	***	1.61	***
	(0.06)		(.09)	
Instructor B	-.02		.98	
	(0.07)		(.07)	
Constant	0.19			
	(0.07)			
Student	0.21			
	(0.04)			
Time	0.49	***		
	(0.05)			

Note: N = 15,242 (# items for all students across all times). The model is a mixed-effects logistic regression. Time (8) and Student are nested variables at level-2 and level-3 respectively. Random effects are not reported when transformed to odds ratio. (**p < .05; ***p < .01).

and time (level 3) were also included in the model. The random effects simply account for the differences in abilities between students, and for each students' change in ability across time. By not assuming all students are equally able to answer MCQs and that each student's ability might change as their experience with academic English increases, we are able to better estimate the unique effect of unpacking each question. The logit odds are reported in Table 2. However, a somewhat more intuitive interpretation of the results is suggested by the odds ratio described above, which is also reported in Table 2 and is what we use to discuss our results.

As a preliminary analysis, we included an indicator variable about which test format students took (coded 1 if Test A and 0 if Test B) to investigate if the order of seeing three unpacked questions first in Test A rather than two in Test B gave students an “unfair” advantage in being able to then answer the paired question on the second half of the exam. No significant effects were found, which confirmed our hypothesis that significant learning effects between exam parts would not occur, and justified further analysis of both exams jointly. Likewise, given the non-significant finding, we do not include exam type in subsequent analyses or results.

3. Results and discussion

Overall, results confirm our initial hypothesis that unpacking increases students’ ability to demonstrate their knowledge by answering more MCQs correctly. On average, students were 8% more likely to answer an unpacked question correctly when compared to complex questions. These findings show that these EAL students are more likely to score higher on assessment questions that have been unpacked by EAP instructors in the effort to reduce linguistic complexity, indicating that unpacking increases the ability for these EAL students to demonstrate their knowledge of disciplinary content.

The question might arise as to what constitutes a meaningful change in student test performance. Some might question whether or not an 8% improvement in test performance is worth the additional resources required to unpack test items. In the context of our program, instructors regularly cite a discrepancy between the course averages of the students in our pathway program and their direct entry peers (i.e., students admitted directly into first-year studies without the need to complete any additional language requirements). Such an improvement in test performance offers one possible way of bridging this gap in performance, and is thus highly valued in our context.

It is important to note that the focus of this study is on unpacking language for EAL students using a combination of linguistic interventions. That is, we did not isolate and evaluate whether unpacking nominal groups, increasing the use of more frequent vocabulary or deciphering cultural knowledge was individually responsible for increased comprehension, nor did we determine which of these variables was most helpful to students. This exploratory study attempts to show how

unpacking language using a combination of these three interventions can be beneficial for student comprehension. Our rationale here was to first determine if a combination of these interventions would increase performance, which can act as a precursor to future research that seeks to determine whether one individual intervention leads to significant improvement on test scores as well. However, as noted above, when unpacking language, it is sometimes difficult to unpack a nominal group without changing vocabulary, posing some challenges in isolating these interventions as separate variables.

Surprisingly, students in instructor A's class were also more likely to answer questions correctly for all test items included in the study, regardless of whether or not they were unpacked. Instructor A taught Sociology, whereas instructors B and C (the referent) were both Psychology instructors. While the reason for this difference is unclear, there is some possibility that the different test scores in instructor A to B and C's class reflects students' ability to comprehend the language of either discipline. Hyland (2004) explains that there are considerable differences in the way that language is used across disciplines. There is some possibility that these two disciplines vary in their degree of technical and lexically dense language, potentially impacting student comprehension. However, further research would be necessary to determine such a possibility. Finally, there was a significant amount of variance explained at the student and time levels, which merely reflects that students have different abilities and that their ability to answer course-specific MCQs changes across time due to either overall test difficulty, change in student ability, or other test performance factors like anxiety.

4. Conclusions and implications

Students' increased performance on unpacked MCQs highlights the importance of unpacking dense academic language used in assessments to provide linguistic space for novice students to better understand the complexity commonly found in academic assessments. This study has several pedagogical implications for educators in both EAP and non-EAP contexts, as well as settings in which instructors collaborate across these disciplines. Two overarching questions arise from these findings. Should instructors of disciplinary courses such as Psychology and Sociology pay more attention to unpacking test items, and by extension, written academic language for novice students early in their postsecondary careers? Or, as students will inevitably encounter complex test questions throughout their academic careers, should EAP instruction focus its efforts on helping students decode lexically dense language and familiarizing them with more infrequent vocabulary, along with equipping them with greater cultural knowledge of local contexts? Of course, these two considerations are not mutually exclusive as can be seen in the approach our program adopts. Instructors must both design MCQ assessments that contain language "appropriate for the students and for the learning objectives" (Parkes & Zimmaro, 2016, p. 24) while systematically supporting students over the course of a curriculum in a way that prepares them to produce complex academic written discourse.

Similar to Miller, Mitchell and Pessoa's (2016) suggestion for instructors to provide explicit prompts for written assignments to assist students' genre uptake, students could benefit both from unpacked academic language as well as explicit language instruction surrounding nominal groups, ultimately helping them to not only comprehend and decode complex language in assessments but also increase the complexity of their own written assignments (Lan & Sun, 2019). In line with Pea's (2004) concept of "fading," greater attention could be placed on unpacking language early in a student's academic career, yet as the terms progress, and such instruction is provided, this emphasis on unpacking could be gradually reduced. In essence, we fully agree with Parkes and Zimmaro's (2016) call for a balance between both accommodating and preparing students for the rigours of participating in academic discourse. The use of complex language is essential when it is related to the learning objectives of a course, yet in assessments, "the item-writing rule articulates that the language should be no more complex than it has to be" (p. 24).

In the context of our program, we strive to achieve this balance through carefully sequencing our curriculum in a way that first emphasizes student comprehension and gradually progresses to a central focus on students' own production of complex academic discourse. We begin developing students' reading comprehension skills early in the term by providing them extensive opportunities with unpacking complex academic language. We instil a foundational awareness of the key differences between spoken and written language, with an emphasis on the prevalence of nominal groups in written academic discourse and the associated concept of nominalization. As the term progresses, we place emphasis on academic reading strategies to support students with the challenging academic readings written in a register that is largely unfamiliar to them. As part of this instruction, we teach the different parts of the nominal group, focusing on the differences between pre- and post-modification of the head noun, and in particular, the typical ways that nominal groups are modified.

Following this focus on the comprehension of complex nominal groups, our focus then shifts to student production of nominal groups under the guidance of the instructor. In our lessons, students are provided opportunities to practice packing language into more complex nominal groups, additionally developing awareness of the importance of balancing pre- and post-modification. In many of these classes, the instructor facilitates the joint construction of a relevant short text, eliciting suggestions from students throughout the lesson resulting in a text that is composed collaboratively. Finally, we begin to lead students to independently produce their own complex written academic discourse. We achieve this by emphasizing the process of writing, dividing assignment writing into draft and final submission stages and highlighting the importance of revision. In draft assignments, students are expected to utilize the lessons and course material focused on producing nominal groups and to incorporate these strategies into their writing, along with several other writing strategies. Students receive extensive feedback on these draft assignments, drawing attention to the ways that they have or have not used these strategies successfully. Students are then asked to revise their assignments and resubmit final versions of their written texts. As the term

comes to a close, students are given greater independence, as the final assignment does not include a draft stage in an attempt to gradually withdraw the earlier support provided and assess students' ability to independently produce complex written academic discourse. Through the teaching of nominal groups within the curriculum, we endeavour to equip students with the ability to both comprehend and produce the language required of their disciplines.

While the present study found unpacking to be helpful for students on MCQ assessments, the goal of future research should be to determine to what extent this unpacking supports students' abilities to demonstrate their knowledge depending on their proficiency level as well as the period of time they have been participating in their university courses (i.e., how many terms) so that such unpacking can be systematically withdrawn as student abilities develop. Ultimately, through replication studies and future research, we hope to have a better understanding of how early interventions can be most relevant and appropriate in the beginning of a program when students are refining their academic language skills while also completing credit bearing content courses. Later in the curriculum, this support could be gradually withdrawn as they develop increased proficiency.

As we see a call for increased collaboration between EAP and disciplinary instructors in integrated instructional approaches (Stapleton & Wu, 2015; Wingate, 2018), the outcome of this project could provide additional implications for EAP specialists. In this study, a large amount of time was dedicated to unpacking questions prior to exams, with EAP instructors being responsible for this labour. This presents additional opportunities for EAP specialists to conduct workshops and consult with disciplinary instructors without a background in language education in order to provide them with greater expertise in designing MCQs appropriate to the reading level of their EAL students. Such collaboration between EAP and disciplinary instructors may help disciplinary instructors fulfill the diverse needs of their EAL learners in postsecondary contexts in terms of assessment, curriculum design and instruction more broadly by providing a concrete foundation upon which a deeper collaborative relationship might be built. In addition, there may also be room for potential engagement with textbook publishers, who routinely include MCQ test banks with materials provided to instructors. These collaborations will ideally encourage professional development and greater awareness of EAL student needs across a large range of instructors and institutions.

With the internationalization of education and the many benefits it brings (Andrade, 2006), there is an obligation to provide an increased amount of support for incoming international students to ensure greater fairness and provide them with rich experiences. To further enhance the support provided to these students, there are several limitations of this study that point to possible directions for future research. First, exploring whether certain unpacking strategies are more significant in enhancing students' comprehension of questions could provide a more nuanced understanding of what the priorities of targeted instruction might be. Second, this study treated participants as one group based on the fact that students are not grouped by proficiency in our program. Future research could attempt to track EAL students' specific proficiency level, for example, through reference to a standardized language test, to determine how much unpacking helps students at different proficiencies. Future attention could also be paid to EAL students' previous institutional learning contexts (e.g., international high schools), which may have implications for cultural knowledge and language proficiency. Future longitudinal studies could also measure the effect of this unpacking in upper years of undergraduate studies to determine whether the effect of unpacking decreases as students advance through their degree, as well as the effect of instruction on student ability to comprehend complex language. Closer attention should also be paid to the potential differences in the effect of such unpacking across disciplines. Finally, future studies could attempt to replicate these results in other contexts to increase the generalizability of these findings. As the number of EAL students continues to grow in postsecondary institutions, it will become increasingly important to address such questions to ensure that EAL students are provided with curriculum, including instruction and assessments, which meet their linguistic needs.

Funding

This work was supported by the SoTL Seed program funded by the Institute for the Scholarship of Teaching and Learning (ISoTL) and the Center for Teaching, Learning and Technology (CTLT) at the University of British Columbia.

Declaration of competing interest

None.

References

Abedi, J., & Gándara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice, 25*(4), 36–46.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1–28.

Abedi, J., & Lord, C. (2001). The Language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219–234.

Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal of Research in International Education, 5*(2), 131–154.

Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes, 31*, 58–71.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*, 2–20.
Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5–35.
Duff, P., & Zappa-Hollman, S. (2012). Using popular culture in language teaching. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing.
Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59–84.
Fang, Z., Schleppegrell, M. J., & Cox, B. E. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research, 38*(3), 247–273.
Ferreira, A. A., & Zappa-Hollman, S. (2019). Disciplinary registers in a first-year program: A view from the context of curriculum. In *Language, context and text: The social semiotics forum* (Inaugural Edition).
Hacking, J. F., & Tschirner, E. (2017). The contribution of vocabulary knowledge to reading proficiency: The case of College Russian. *Foreign Language Annals, 50*(3), 500–518.
Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309–333.
Halliday, M. A. K. (1987). Spoken and written modes of meaning. In R. Horowitz, & S. Samuels (Eds.), *Comprehending oral and written language* (pp. 55–82). London: Academic Press.
Halliday, M. A. K. (2007). In J. J. Webster (Ed.), *language and education*. New York, NY: Continuum.
Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. New York: Routledge.
Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). New York: Routledge.
Hsu, C., & Yang, F. (2013). A vocabulary learning tool for L2 undergraduates reading science and technology textbooks. *International Journal of Science Education, 35*(7), 1110–1138.
Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.
Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor: University of Michigan Press.
Hyland, K. (2006). *English for academic purposes: An advanced resource book*. Abingdon, Oxon: Routledge.
Lan, G., & Sun, Y. (2019). A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course. *Journal of English for Academic Purposes, 38*, 14–24.
Lee, D. S. (1997). What teachers can do to relieve problems identified by international students. *New Directions for Teaching and Learning, 70*, 93–100.
Matruglio, E., Maton, K., & Martin, J. R. (2013). Time travel: The role of temporality in enabling semantic waves in secondary school teaching. *Linguistics and Education, 24*(1), 38–49.
McCabe, A., & Gallagher, C. (2008). The role of the nominal group in undergraduate academic writing. In C. Jones, & E. Ventola (Eds.), *language to multimodality: New developments in the study of ideational meaning* (pp. 189–208). London: Equinox.
Miller, R. T., Mitchell, T. D., & Pessoa, S. (2016). Impact of source texts and prompts on students' genre uptake. *Journal of Second Language Writing, 31*, 11–24.
Moore, T. (2007). The 'Process' of Learning: On the use of Halliday's transitivity in academic skills advising. *Arts and Humanities in Higher Education, 6*(1), 50–73.
Parkes, J., & Zimmaro, D. (2016). *Learning and assessing with multiple-choice questions in College classrooms*. New York: Routledge.
Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*(3), 423–451.
Rivera, C., Stansfield, C. W., Scialdone, L., & Sharkey, M. (2000). *An analysis of state policies for the inclusion and accommodation of English language learners in state assessment programs during 1998–1999*. Arlington, VA: George Washington University, Center for Equity and Excellence in Education.
Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33*(2), 149–183.
Stapleton, P., & Wu, Y. (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes, 17*, 12–23.
Storch, N., Morton, J., & Thompson, C. (2016). EAP pedagogy in undergraduate contexts. In K. Hyland, & P. Shaw (Eds.), *The routledge handbook of English for academic purposes* (pp. 477–488). New York, Milton Park: Routledge.
Thompson, G. (2014). *Introducing functional grammar* (3rd ed.). Milton Park, Abingdon, Oxon: Routledge.
Ventola, E. (1996). Packing and unpacking of information in academic texts. In E. Ventola, & A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues*. Amsterdam, Philadelphia: John Benjamins Pub.
Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal, 96*, 576–598.
Wingate, U. (2018). Academic literacy across the curriculum: Towards a collaborative instructional approach. *Language Teaching, 51*(3), 349–364.

Daniel Riccardi teaches EAP at the University of British Columbia, Vantage College, and received the 2019 Vantage One Teaching award. He holds an MA in Language and Literacies Education from the Ontario Institute for Studies in Education, University of Toronto, and has taught EAP in South Korea, Chile and Canada.

Jennifer Lightfoot teaches EAP at the University of British Columbia, Vantage College, holds an MA in English Language Teaching and Applied Linguistics from King's College London, and has taught within Canada and the UK.

Mark Lam is a Lecturer in the Department of Psychology at the University of British Columbia. His research interests include scholarship of teaching and learning, and health psychology.

Katherine Lyon is an Instructor in the Department of Sociology at the University of British Columbia. She is a recipient of the Vantage One Teaching Award and the SAGE Teaching Innovations and Professional Development Award. Her research interests include experiential pedagogies, accessible assessment, and teaching with augmented reality.

Nathan Roberson is a PhD student in Measurement, Evaluation, and Research Methodology at the University of British Columbia. His research interests include: measurement, econometrics, immigration, quality of life, and bilingual education. He holds an M.A. in International Relations from the Central European University in Budapest, Hungary.

Simon Lolliot is an Instructor in the Department of Psychology at the University of British Columbia. His research interests include diversity, intergroup contact, belongingness, and how these relate to success (pedagogical and otherwise) at university; scholarship of teaching and learning and educational leadership.