ETS TOEFL

Quality Beyond Measure.

TOEFL® Research Report
TOEFL–RR-94
ETS Research Report No. RR–21-16

A Comparison of Spoken and Written Language Use in Traditional and Technology-Mediated Learning Environments

Kristopher Kyle

Ann Tai Choe

Masaki Eguchi

Geoff LaFlair

Nicole Ziegler

December 2021

The *TOEFL*® test is the world's most widely respected English language assessment, used for admissions purposes in more than 130 countries including Australia, Canada, New Zealand, the United Kingdom, and the United States. Since its initial launch in 1964, the TOEFL test has undergone several major revisions motivated by advances in theories of language ability and changes in English teaching practices. The most recent revision, the *TOEFL iBT*® test, contains a number of innovative design features, including integrated tasks that engage multiple skills to simulate language use in academic settings and test materials that reflect the reading, listening, speaking, and writing demands of real-world academic environments. In addition to the TOEFL iBT, the TOEFL Family of Assessments has expanded to provide high-quality English proficiency assessments for a variety of academic uses and contexts. The TOEFL Young Students Series (YSS) features the *TOEFL*® *Primary*™ and *TOEFL Junior*® tests, designed to help teachers and learners of English in school settings. The *TOEFL ITP*® Assessment Series offers colleges, universities, and others an affordable test for placement and progress monitoring within English programs.

Since the 1970s, the TOEFL tests have had a rigorous, productive, and far-ranging research program. ETS has made the establishment of a strong research base a consistent feature of the development and evolution of the TOEFL tests, because only through a rigorous program of research can a testing company demonstrate its forward-looking vision and substantiate claims about what test takers know or can do based on their test scores. In addition to the 20-30 TOEFL-related research projects conducted by ETS Research & Development staff each year, the TOEFL Committee of Examiners (COE), composed of distinguished language-learning and testing experts from the academic community, funds an annual program of research supporting the TOEFL family of assessments, including projects carried out by external researchers from all over the world.

To date, hundreds of studies on the TOEFL tests have been published in refereed academic journals and books. In addition, more than 300 peer-reviewed reports about TOEFL research have been published by ETS. These publications have appeared in several different series historically: TOEFL Monographs, TOEFL Technical Reports, TOEFL iBT Research Reports, and TOEFL Junior Research Reports. It is the purpose of the current TOEFL Research Report Series to serve as the primary venue for all ETS publications on research conducted in relation to all members of the TOEFL Family of Assessments.

To obtain more information about the TOEFL programs and services, use one of the following:

E-mail: toefl@ets.org **Web site:** www.ets.org/toefl

TOEFL Research Report Series and ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

A Comparison of Spoken and Written Language Use in Traditional and Technology-Mediated Learning Environments

Kristopher Kyle[1,2], Ann Tai Choe[3], Masaki Eguchi[1], Geoff LaFlair[3], & Nicole Ziegler[3]

1 University of Oregon, Eugene, OR
2 Yonsei University, Seoul, South Korea
3 University of Hawai'i at Mānoa, Mānoa, HI

A key piece of a validity argument for a language assessment tool is clear overlap between assessment tasks and the target language use (TLU) domain (i.e., the domain description inference). The TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) corpus, which represents a variety of academic registers and disciplines in traditional learning environments (e.g., lectures, office hours, textbooks, course packs), has served as an important foundation for the *TOEFL iBT*® test's domain description inference for more than 15 years. There are, however, signs that the characteristics of the registers that students encounter may be changing. Increasingly, typical university courses include technology-mediated learning environments (TMLEs), such as those represented by course management software and other online educational tools. To ensure that the characteristics of TOEFL iBT test tasks continue to align with the TLU domain, it is important to analyze the registers that are typically encountered in TMLEs. In this study, we address this issue by collecting a relatively large (4.5 million words) corpus of spoken and written TMLE registers across the six primary disciplines represented in T2K-SWAL. This corpus was subsequently tagged for a wide variety of linguistic features, and a multidimensional analysis was conducted to compare and contrast written and spoken language in TMLE and T2K-SWAL. The results indicate that although some similarities exist across spoken and written texts in traditional learning environments and TMLEs, language use also differs across learning environments (and modes) with regard to key linguistic dimensions.

Keywords Corpus linguistics; validity; academic English; technology-mediated learning environments

doi:10.1002/ets2.12329

An important aspect of a validity argument for a language assessment tool such as the *TOEFL iBT*® test is a demonstrated alignment between the linguistic demands of the target language use (TLU) domain and the assessment tasks (Chapelle et al., 2008). Such alignment provides evidence for key inferences, such as the domain description inference and the extrapolation inference. Corpus analyses are well suited to generating evidence for the domain description and extrapolation inferences (Biber et al., 2004), but researchers are constrained by the degree to which available corpora represent the target domains. Currently a number of corpora represent various types of language that university students encounter and/or produce in traditional academic settings, such as the British Academic Written English (BAWE) corpus (Alsop & Nesi, 2009), the Michigan Corpus of Upper-Level Student Papers (MiCUSP) corpus (Römer & O'Donnell, 2011), and TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL; Biber et al., 2004). Previous seminal studies (Biber, 2006; Biber et al., 2004) examined the linguistic features of traditional learning environments through the collection and multidimensional analysis (MDA) of the T2K-SWAL corpus. Biber et al. (2004) provided a linguistic foundation for the TOEFL iBT test validity argument (Chapelle et al., 2008) by describing the linguistic features of the target domain (and the degree to which those features varied by mode and register).

Although T2K-SWAL remains an important corpus that represents the language of traditional university learning environments in the United States, a typical university experience is now increasingly supported by technology-mediated learning environments (TMLEs; e.g., Jacoby, 2014; Means et al., 2013), which are not represented in extant academic corpora. Consequently, it is currently unclear the degree to which (and how) spoken and written language in TMLEs differs from that in more traditional academic learning environments. If spoken and/or written TMLE language differs from traditional environments with regard to the actual language features represented and/or the relative proportion of shared features (e.g., affecting reading or listening difficulty), it may be important to consider these differences in test development. The current study attempted to address this research gap by collecting a corpus of spoken and written TMLE texts.

Corresponding author: K. Kyle, E-mail: kristopherkyle1@gmail.com

The linguistic features of these texts were then compared and contrasted with the linguistic features of spoken and written texts representing traditional learning environments (the T2K-SWAL corpus) using a MDA to determine the degree to which (and in what ways) the language in TMLEs differed from that in traditional learning environments.

Literature Review

Corpus-based analyses can be used to highlight the linguistic features that are typical of a particular domain and to highlight how situational differences between domains (e.g., mode, register) can affect the particular linguistic features and/or the distribution of features that occur in a domain. Biber et al. (2004), for example, explored the linguistic features of a number of academic written and spoken registers using MDA (see the Literary Analyses section for a description of MDA). Among other findings, their results indicate that spoken academic texts share a set of features, such as the use of interactive language and personal involvement (e.g., the use of *I*, *we*, and *you*) and reduced language forms (e.g., contractions and *that*-deletion). Academic written texts, by contrast, tend to be characterized by complex noun phrases and other linguistic structures that can be used to increase informational density (e.g., nominalizations). In addition, they found that spoken and written academic language also varies by register. Among spoken texts, for example, service encounters had the highest proportion of language use related to interaction and personal involvement (among others), while classroom teaching had fewer of these features (but still more than any written registers). A description of the linguistic features of a particular domain (e.g., academic language) can be particularly helpful for designing language assessments and providing evidence for key validity argument inferences, such as the domain description inference and the extrapolation inference (Chapelle et al., 2008).

The domain description inference presumes that assessment tasks are representative of the target domain. Biber et al.'s (2004) corpus-based investigation and description of the linguistic features of academic language use represents an important aspect of the TOEFL iBT test's support for the domain description inference. Biber et al.'s research revealed that mode and register are two key predictors of linguistic variation and provided an empirically based description of the linguistic features of various academic registers (across modes). Descriptions of the typical linguistic features of particular academic registers can be and have been used in test development to help support the domain description inference by demonstrating that the linguistic features of reading and listening passages in an assessment align with those in the target domain. Biber et al.'s results, for example, were used to develop an evaluation tool for reading passages being considered for the TOEFL iBT test (Sheehan et al., 2008, 2010).

Relatedly, corpus evidence can also be used to support the extrapolation inference, which presumes that test users can make inferences about test-taker language use abilities in target domains (such as academic reading, listening, speaking, and/or writing) from their performance on the test. In the context of language proficiency assessments, this means that the linguistic skills (among other skills) required for the successful completion of a task should match those typically encountered in the TLU domain. If target domain-irrelevant linguistic features contribute to the difficulty of an assessment task, then the support for the extrapolation inference is weakened. Similarly, if key linguistic features of a particular domain are not required for the successful completion of a task, support for the extrapolation inference is also weakened. Conversely, if the use of key linguistic features for a domain is necessary for the successful completion of a task, support for the extrapolation inference is warranted (see, e.g., Brooks & Swain, 2014; LaFlair & Staples, 2017; Sheehan et al., 2010).

Technology-Mediated Learning Environments

Over the last decade, TMLEs, such as online and blended classrooms, have become increasingly prevalent in higher education (Jacoby, 2014; Means et al., 2013). In some cases, the texts encountered and produced by students in TMLEs mirror those encountered and produced in traditional learning contexts (Naidu, 2013). However, in many cases, TMLEs include a wide range of texts that leverage technological affordances not found in traditional learning environments. TMLEs, for example, may make use of technology such as learning management systems, simulations, multiuser games, wikis, blogs, synchronous and asynchronous computer-mediated communication, and social media (Means et al., 2013). Although linguistic analyses have been conducted on some TMLE texts (e.g., Crossley et al., 2016), these analyses have focused on predicting student performance, not on describing the linguistic features of such texts. Little, therefore, is known regarding the linguistic features of texts that represent TMLEs or how those linguistic features may differ from or be similar to more traditional registers of academic language. If there are substantial differences in the linguistic features used and/or the distribution of those features between TMLEs and traditional learning environments, it is possible that the support for the domain description inference and the extrapolation inference for academic language proficiency assessments like the TOEFL iBT test may need to be bolstered.

To model the linguistic demands of the university experience, it is important to have a representative sample of the registers that are part of that experience. Extant corpora, such as BAWE, MiCUSP, the Oxford corpus, and T2K-SWAL, arguably together compose a representative sample of the language encountered and produced in traditional university settings. Increasingly, however, the university experience includes TMLEs, which are not represented in the aforementioned corpora. This lack of representation suggests that a supplemental corpus may be needed to model the types of language used in the university experience.

Linguistic Analyses

LaFlair and Staples et al. (2017) provided empirical support for the use of corpus-based register analysis (Biber & Conrad, 2014) as a tool for evaluating linguistic evidence that supports, or refutes, the key inferences for assessments of productive language skills (speaking and writing). They demonstrated that Biber and Conrad's (2014) framework for corpus-based register analysis, which is based on earlier seminal work by Biber (1988), aligns with Bachman and Palmer's (1996, 2010) framework for conducting TLU domain analysis. Both Biber and Conrad's corpus-based register analysis and Bachman and Palmer's framework for TLU domain analysis are grounded in theories of communicative competence (see discussion in LaFlair & Staples, 2017) but were developed in different fields of applied linguistics (socio/corpus linguistics and language assessment, respectively) and have different purposes. The purpose of corpus-based register analysis is to identify the pervasive communicative linguistic features of various registers, contrast those results across registers, and explain the differences in linguistic features across registers in light of the differences in their situational characteristics, which include (but are not limited to) relationships between participants, communicative purpose, topic, and channel. Bachman and Palmer's TLU domain analysis uses these same characteristics for a different purpose. Instead of seeking to describe the use of language across registers, the purpose of carrying out a TLU domain analysis is to identify characteristics that may affect the language elicited by a test task and use those characteristics in the design of the test task to attempt to elicit language that is similar to the language used in the intended target domain (Bachman, 1990). Furthermore, TLU domain analysis and corpus-based register analysis are both grounded in theories of communicative competence that highlight the sensitivity of language use to differences in situational characteristics (Canale & Swain, 1980; Hymes, 1974).

The predominant method for the linguistic analysis and comparison of registers is MDA (Biber, 1988; Biber et al., 2004; Biber & Conrad, 2014), which involves four main steps after creating, or identifying, representative corpora of registers under study. The first step is to identify the relevant situational characteristics of the domain that is under investigation. In academic domains, for example, there are important distinctions between spoken texts (e.g., lectures) and written texts (e.g., articles and textbook chapters). There are also various settings in which texts can be encountered that may affect the linguistic features of those texts (e.g., formality, social distance). The second step is to annotate the texts in a corpus for relevant linguistic features. Although a wide range of linguistic features can be annotated, the features identified by Biber and colleagues (e.g., Biber, 1988; Biber et al., 2004) are a common starting point. In most cases, an automatic tagging system (such as the Biber Tagger; Biber, 1988) is used to efficiently annotate the corpus/corpora. The third step is to conduct an exploratory factor analysis (EFA) on the set of linguistic features found in all registers and interpreting the factors to understand the underlying linguistic dimensions in respect to their communicative functions. This step is aided by the initial analysis of the situational characteristics of the domain (outlined in the first step). For example, in most MDA analyses, the first dimension tends to highlight differences between written and spoken language (e.g., the use of personal pronouns and contractions vs. the use of complex noun phrases). The final step is to compare the registers in the corpus with regard to each dimension by calculating a dimension score for each text and comparing the mean dimension score for each register of interest. Biber et al. (2004), for example, tagged the T2K-SWAL corpus for 159 features related to words, multiword units (lexical bundles), and lexicogrammatical features. The EFA resulted in the identification of four dimensions (e.g., oral vs. literate discourse). Different registers were then compared with regard to scores for each dimension. Service encounters, office hours, labs, and study groups, for example, scored very high on the oral versus literate discourse dimension (indicating a high incidence of features such as contractions and personal pronouns and a low incidence of features such as complex noun phrases), while institutional writing scored very low on this dimension (indicating a low incidence of features such as personal pronouns and contractions and a high incidence of features such as complex noun phrases), which highlights a consistent difference in the lexico-grammatical features of spoken versus written language.

In this study, we build on previous work related to academic language through the collection of a corpus of spoken and written language use from TMLEs. We then conduct a new MDA using spoken and written texts from both traditional learning environments (represented by T2K-SWAL) and TMLEs with the goal of comparing and contrasting the features of spoken and written language use within and across learning environments.

Method

This study explores the similarities and differences in spoken and written language use across traditional learning environments and TMLEs. First, a corpus of spoken and written language used in TMLEs is collected and compiled. Then, an MDA is conducted to compare language use across modes and learning environments. This project is guided by the following research question: How (dis)similar are the features of spoken and written language use within and across traditional learning environments and TMLEs?

Corpora

TOEFL 2000 Spoken and Written Academic Language Corpus

The T2K-SWAL corpus (Biber et al., 2004) consists of approximately 2.7 million words that are representative of the kinds of spoken and written university registers encountered across various traditional academic settings and academic disciplines (for further details, see Biber et al., 2004). The T2K-SWAL corpus was chosen to represent traditional academic settings in this study because it has been used in a number of seminal and TOEFL iBT test–focused studies (e.g., Biber, 2006; Biber et al., 2004).

The T2K-SWAL corpus was collected from participants who were recruited to document naturally occurring discourse as they engaged in a range of academic activities. Both the spoken and written portions of the corpus were, by and large, sampled from six major disciplines: business, education, engineering, humanities, natural sciences, and social sciences. The spoken corpus (1.7 million words) was collected at four public universities in the United States. Students at these universities were recruited to tape-record class lectures, study group sessions, and other academic-related discussions. Faculty members and university staff at the universities were also recruited to tape-record office hours and academic service encounters (e.g., university bookstores, business services, libraries reference desks), respectively, which were subsequently transcribed. The written corpus (1 million words) is made up of textbooks, course packs (e.g., lecture notes, study guides, assignment descriptions), and any miscellaneous institutional writings (e.g., program brochures and university catalogs) sampled from the six major disciplines. Table 1 breaks down the T2K-SWAL corpus by mode and text type; Table 2 breaks down the corpus by mode and discipline.

Technology-Mediated Learning Environment Corpus

To address the current lack of academic corpora that represent TMLEs, we designed and constructed the TMLE corpus (~4.5 million words) to be representative of the various types of spoken and written registers that university students encounter and produce in TMLEs (e.g., instructional videos, blog posts, online discussion forums).

Corpus Collection

The corpus collection process involved two main stages. First, we contacted instructors of undergraduate courses in our target disciplines (business, education, engineering, humanities, natural sciences, and social sciences) at three large public universities in the United States via e-mail. We asked the instructors to provide any materials that students in their class(es) encounter via TMLEs (e.g., via course management software). Additionally, we asked instructors to send a recruitment message to their students asking for additional texts produced by students in or for TMLEs. Texts were collected via an online interface that also collected metadata (e.g., course name, discipline, subdiscipline, assignment type). Participants who agreed to share texts received a $10 gift card after submitting at least five texts.

Table 1 Breakdown of TOEFL 2000 Spoken and Written Academic Language Corpus by Mode and Text Type

Mode	Text type	No. texts	No. words
Spoken	Classroom management talk	38	35,669
	Lab	17	90,792
	Lecture	177	1,279,659
	Office hours	11	49,472
	Service encounters	22	105,810
	Study group	25	157,349
Subtotal		290	1,718,751
Written	Course management	21	52,791
	Course packs	27	108,578
	Other institutional writing	37	153,061
	Textbooks	87	764,449
Subtotal		172	1,078,879
Total		462	2,797,630

Note. A small number of TOEFL 2000 Spoken and Written Academic Language texts had been corrupted and were not used in this study. The numbers in this table represent the version of the corpus that was used in this study. Adapted from Biber et al. (2004).

Table 2 Breakdown of TOEFL 2000 Spoken and Written Academic Language Corpus by Mode and Discipline

Mode	Discipline	No. texts	No. words
Spoken	Business	52	296,262
	Education	26	173,930
	Engineering	40	209,279
	Humanities	46	314,110
	Natural science	52	272,669
	Service encounters	22	105,810
	Social science	52	346,691
Subtotal		290	1,718,751
Written	Business	18	126,925
	Education	15	89,248
	Engineering	20	105,748
	Humanities	29	193,609
	Natural science	22	136,801
	Social science	38	302,845
	Other	30	123,703
Subtotal		172	1,078,879
Total		462	2,797,630

Note. A small number of TOEFL 2000 Spoken and Written Academic Language texts had been corrupted and were not used in this study. The numbers in this table represent the version of the corpus that was used in this study.

Second, to supplement the size of the corpus and balance the number of spoken and written registers across six major disciplines (business, education, engineering, humanities, natural sciences, and social sciences), we also sampled texts from a popular publicly available massive open online course (MOOC) platform.

The spoken corpus (2.2 million words) includes online classroom management talk (e.g., course announcements, discussions, and assignments) and instructional videos. The written corpus (2.2 million words) contains similar text types as the spoken corpus in addition to any reading materials (e.g., web pages, PDF files, and syllabi), quizzes, and lecture slides that were posted in the TMLEs. All texts were transcribed into Extensible Markup Language (XML) data files by trained transcriptionists following a standard protocol. For example, in instructional videos involving multiple speakers, the speakers were identified either by their names or as Speaker 1, Speaker 2, and so on, followed by a colon to indicate their turn. For videos that included displayed text on-screen, transcribers inserted comments using a pair of square brackets (e.g., [Screen: displayed text]). Filler words and repairs by the speakers were eliminated from the transcriptions. For written materials, such as articles, forums, and slides, paragraph and sentence boundaries were delimited by a period. All spoken

Table 3 Breakdown of Technology-Mediated Learning Environment Corpus by Mode, Text Type, and Collection Context

Mode	Text type	Collection context	No. texts	No. words
Spoken	Classroom management talk	MOOC	27	11, 591
	Instructional video	MOOC	2,135	2,118,667
		Public university	82	155,084
Subtotal			2,244	2,285,342
Written	Announcements/discussions	MOOC	232	37,422
		Public university	171	53,805
	Assignment description	MOOC	230	87,507
		Public university	160	119,741
	Instructional reading	MOOC	514	700,508
		Public university	149	799,441
	Quiz	MOOC	256	40,958
		Public university	50	30,873
	Slides	MOOC	27	23,296
		Public university	113	128,092
	Syllabus	MOOC	37	27,282
		Public university	142	214,772
Subtotal			2,081	2,263,697
Total			4,325	4,549,039

Note. MOOC = massive open online course.

texts were either automatically transcribed and then hand checked for accuracy and consistency through postediting by the research assistants or manually transcribed by trained research assistants. Table 3 breaks down the TMLE corpus by mode, text type, and collection context; Table 4 breaks down the TMLE corpus by mode, discipline, and collection context.

Linguistic Features

The linguistic features that were considered include a selection of commonly used and newer complexity measures (Kyle & Crossley, 2018; Lu, 2011) in addition to the grammatical and lexicogrammatical features investigated by Biber et al. (2004) and Biber (2006). All features were tagged using the Tool for the Analysis of Syntactic Sophistication and Complexity (Kyle, 2016) version 2.0, which makes use of the natural language processing tool Spacy (Explosion AI, 2018). These are described in the following sections.

Complexity Features

Tables 5–7 outline the complexity features used.

Grammatical and Lexicogrammatical Features

Tables 8–17 compose a description of the grammatical and lexicogrammatical features considered, which (with a few minor exceptions) align with the grammatical and lexicogrammatical features described in Biber et al. (2004). Note that complete lists of the words included in each semantic category can also be found in Biber et al.

Statistical Analyses

To identify register variation in TMLEs, a new MDA was conducted using all of the texts included in both the TMLE corpus and T2K-SWAL. In general, a new MDA is justified when researchers expect that the texts included in the corpus contain situationally different registers from the corpus used to conduct the previous MDA (for detailed discussion, see Biber, 2006; Biber et al., 2004). Because it was expected that the TMLE corpus would possibly introduce new situational contexts of language use, a new MDA was deemed appropriate. All the statistical analyses were conducted through the *R* statistical package (R Development Core Team, 2014).

Table 4 Breakdown of Technology-Mediated Learning Environment Corpus by Mode, Discipline, and Collection Context

Mode	Discipline	Collection context	No. texts	No. words
Spoken	Business	MOOC	645	657,720
	Education	MOOC	178	232,287
		Public university	32	42,586
	Engineering	MOOC	396	279,414
		Public university	14	22,020
	Humanities	MOOC	398	346,572
		Public university	7	10,287
	Natural science	MOOC	188	215,340
		Public university	22	69,338
	Social science	MOOC	357	398,925
		Public university	7	10,853
Subtotal			2,244	2,285,342
Written	Business	MOOC	445	146,680
		Public university	111	425,201
	Education	MOOC	118	80,800
		Public university	162	198,530
	Engineering	MOOC	241	234,361
		Public university	63	90,168
	Humanities	MOOC	59	44,120
		Public university	135	282,528
	Natural science	MOOC	192	99,748
		Public university	151	163,519
	Social science	MOOC	241	311,264
		Public university	163	186,778
Subtotal			2,081	2,263,697
Total			4,325	4,549,039

Note. MOOC = massive open online course.

Table 5 Classic Measures of Syntactic Complexity, Lexical Sophistication, and Lexical Diversity

Feature	Description/example
Mean length of clause (MLC)	Number of finite clauses/number of total words
Mean length of T-unit (MLTU)	Number of independent finite clauses/number of total words
Dependent clauses per clause (DC/C)	Number of finite dependent clauses/number of finite clauses
Word length	Mean number of characters per word
Moving-average type–token ratio (MATTR)	Average type–token ratio based on a 50-word moving window

Table 6 Fine-Grained Clausal Complexity Measures

Feature	Description/example
Dependents per clause	Mean number of direct dependents per finite verb
Clausal complements per clause	Mean number of clausal complements per finite clause
Relative clauses per clause	Mean number of relative clauses per finite clause
Nonfinite clause proportion	Proportion of all clauses that are nonfinite clauses

Multidimensional Analysis

According to Biber (1988), MDA comprises a series of quantitative analyses: conducting an EFA, calculating dimensional scores based on the final EFA solution, and comparing the differences of mean dimensional scores across different text categories. To maximize the comparability of our MDA results of TMLE discourse to previous MDAs on traditional classroom discourse, we decided to replicate the statistical procedures of previous prominent studies (i.e., Biber, 2006; Biber et al., 2004). These statistical procedures are outlined in the following paragraphs. One terminological note is necessary here; we use *factor* in the statistical sense to denote the latent factor we extract through the common factor model, whereas

Table 7 Fine-Grained Phrasal Complexity Measures

Feature	Description/example
Dependents per nominal	Mean number of direct dependents per common noun (grammatical dependents based on the dependency parse)
Relative clauses per nominal	Mean number of relative clause modifiers per common noun
Adjective modifiers per nominal	Mean number of adjective modifiers per common noun
Determiners per nominal	Mean number of determiners per common noun
Prepositional phrases per nominal	Mean number of prepositional phrase modifiers per common noun
Possessives per nominal	Mean number of possessives per common noun
Coordinated nominals per nominal	Mean number of coordinators per common noun

Table 8 Pronouns and Pro-verbs

Feature	Description/example
First person pronouns	*I*, *we*, *our*, *us*, *my*, *me*, *ourselves*, *myself*
Second person pronouns	*you*, *your*, *yourself*, *ya*, *thy*, *thee*, *thine*
Third person pronouns (excluding *it*)	*he*, *she*, *they*, *their*, *his*, *them*, *her*, *him*, *themselves*, *himself*, *herself*
Pronoun *it*	
Demonstrative pronouns	*this*, *that*, *these*, *those* as pronouns; *So how can we check* ***this*** *with data?*
Indefinite pronouns	*everything*, *someone*, *anybody*, *nobody*
Pro-verb *do*	*You are not required to view these, but if you* ***do****, you may want to try replicating them in the Distribution Simulator*

Note. Bold indicates the target item (e.g., adjective). In cases where examples are more than one word (e.g., that clauses controlled by a verb), the entire “that” clause is bolded, and the controlling word is underlined.

Table 9 Reduced Forms and Dispreferred Structures

Feature	Description/example
Contractions	*'m*, *'ll*, *n't*, *'re*, *'s*
Complementizer *that*-deletion	*I think* ***[0]*** *there's two parts to that answer*
Split auxiliaries	*I* ***will <u>often not</u> use*** *a semicolon just for didactic purposes*

Note. Contraction *'s* is disambiguated from genitive *'s*. Bold indicates the target item (e.g., adjective). In cases where examples are more than one word (e.g., that clauses controlled by a verb), the entire “that” clause is bolded, and the controlling word is underlined.

Table 10 Prepositional Phrases and Coordination

Feature	Description/example
Prepositional phrases	All prepositional phrases
Phrasal coordination	*In Section 4.3, you will* ***define <u>and</u> call functions*** *to perform various operations*
Independent clause coordination	***So Xena goes to the auction, <u>and</u> she bids $20*** *before Bob can get his paddle up*

Note. Bold indicates the target item (e.g., adjective). In cases where examples are more than one word (e.g., that clauses controlled by a verb), the entire "that" clause is bolded, and the controlling word is underlined.

Table 11 Nouns

Feature	Description/example
All nouns	All common and proper nouns
Nominalization	Derived nouns with endings listed in the Longman Grammar (e.g., *-ent*, *-er*, *-tion*, *-ship*). A stoplist was used to ignore frequent nouns that have the form of nominalized endings but are not derived forms (e.g., *percent*, *chapter*, *number*)
Animate noun	*author*, *professor*, *researcher*, *student*
Cognitive noun	*analysis*, *concern*, *conclusion*, *recognition*
Concrete noun	*computer*, *instrument*, *muscle*, *solid*
Technical/concrete noun	*bacteria*, *chromosome*, *software*, *virus*
Quantity noun	*date*, *frequency*, *rate*, *semester*
Place noun	*cave*, *estuary*, *desert*, *museum*
Group/institution noun	*committee*, *firm*, *laboratory*, *university*
Abstract/process noun	*activity*, *contribution*, *explanation*, *strategy*

Table 12 Verbs

Feature	Description/example
Past tense	Verbs tagged as past tense (e.g., *ran*, *was*, *were*, *studied*)
Perfect aspect verbs	*We **had learned** to partition the matrices*
Nonpast tense	Verbs not tagged as past tense or perfect aspect
Agentless passives	*Only recently is its value for disaster reduction **being examined***
By passives	*They were sort of postcolonial cities that **had been designed by colonial administrators** for colonial administrators*
Possibility modals	*can*, *may*, *might*, *could*
Necessity modals	*ought*, *must*, *should*
Predictive modals	*will*, *would*, *shall*
Be as main verb	*to be* as a copular verb
Activity verb	*give*, *hold*, *make*, *show*
Communication verb	*argue*, *describe*, *say*, *write*
Mental verb	*consider*, *expect*, *know*, *think*
Causative verb	*affect*, *enable*, *influence*, *require*
Occurrence verb	*become*, *develop*, *increase*, *rise*
Existence verb	*appear*, *exist*, *possess*, *represent*
Aspectual verb	*begin*, *complete*, *continue*, *start*
Intransitive activity phrasal verb	*come on*, *come over*, *go ahead*, *sit down*
Transitive activity phrasal verb	*pick up*, *look up*, *make up*, *set up*
Transitive mental phrasal verb	*find out*, *give up*
Transitive communication phrasal verb	*point out*
Intransitive occurrence phrasal verb	*come off* , *run out*
Copular phrasal verb	*turn out*
Aspectual phrasal verb	*go on*

Note. Bold indicates the target item (e.g., adjective). In cases where examples are more than one word (e.g., that clauses controlled by a verb), the entire "that" clause is bolded, and the controlling word is underlined.

Table 13 Adjectives

Feature	Description/example
Attributive adjectives	*Furthermore, the trip will end up costing you more because of the **high** interest rate credit card companies charge*
Size-attributive adjectives	*big*, *high*, *little*, *small*
Time-attributive adjectives	*new*, *old*, *young*
Color-attributive adjectives	*white*, *dark*, *red*
Evaluative-attributive adjectives	*best*, *important*, *right*, *simple*
Relational-attributive adjectives	*basic*, *common*, *general*, *various*
Topical-attributive adjectives	*economic*, *international*, *political*, *public*
Predicative adjectives	*You are **welcome** to connect with other students through our discussion forum to organize meetups or connect virtually*

Note. Bold indicates the target item (e.g., adjective). In cases where examples are more than one word (e.g., that clauses controlled by a verb), the entire “that” clause is bolded, and the controlling word is underlined.

Table 14 Adverbs and Adverbials

Feature	Description/example
Place adverbials	*above*, *beneath*, *hereabouts*, *outside*
Time adverbials	*again*, *early*, *presently*, *initially*
Conjuncts	*alternatively*, *furthermore*, *namely*, *therefore*
Downtoners	*almost*, *enough*, *fairly*, *virtually*
Hedges	*kind of*, *sort of*, *almost*, *maybe*
Amplifiers	*absolutely*, *obviously*, *perfectly*, *sufficiently*
Emphatics	*just*, *really*, *so*, *real*
Discourse particles	Sentence-initial *well*, *now*, *anyway*, *anyhow*
Causative adverbial subordinator	*because*
Conditional adverbial subordinator	*if*, *unless*
Other adverbial subordinator	*since*, *while*, *whereas*
Other adverbs	Adverbs not included in preceding types

Table 15 Nominal Postmodifying Clauses

Feature	Description/example
That relative clauses	*Histograms sacrifice just a bit of information to produce plots* ***that are much easier to interpret***
WH-relative clauses	
WH relatives on object position	*But for now, I think we'll move on to the next question,* ***which I also found super interesting***
WH relatives on subject position	*And one of them emits neutrons,* ***which are heavy uncharged particles***
WH relatives with fronted preposition	*And he observed that a voluntary attempt to overcome unnecessary obstacles in* ***which the outcome is uncertain*** *is a game*
Past participial postnominal (reduced relative) clauses	*This little guy,* ***named in honor of the mathematician Euler****, is called the Eulerian acceleration component*

Note. Bold indicates the target item (e.g., adjective). In cases where examples are more than one word (e.g., that clauses controlled by a verb), the entire "that" clause is bolded, and the controlling word is underlined.

Table 16 *That* Complement Clauses

Feature	Description/example
That complement clauses (all)	
That clauses controlled by a verb	*At the same time, I also* ***<u>heard</u> that some people were feared***
Nonfactive verb	*argue*, *claim*, *recommend*, *stress*
Attitudinal verb	*concede*, *feel*, *hope*, *prefer*
Factive verb	*calculate*, *demonstrate*, *observe*, *realize*
Likelihood verb	*appear*, *estimate*, *predict*, *think*
That clauses controlled by an adjective (all)	*You can see how* ***<u>useful</u> that can be***
Attitudinal adjectives	*amused*, *appropriate*, *crucial*, *essential*
Likelihood adjectives	*likely*, *possible*, *probable*, *unlikely*
That clauses controlled by a noun (all)	*And also, a big change from other local currencies was to persuade the local* ***<u>government</u> that they could accept taxes from both individuals and from businesses paid in the local currency***
Nonfactive noun	*comment*, *proposition*, *report*, *requirement*
Attitudinal noun	*hope*, *reason*, *view*, *thought*
Factive noun	*assertion*, *conclusion*, *result*, *statement*
Likelihood noun	*assumption*, *belief*, *idea*, *notion*

Note. Bold indicates the target item (e.g., adjective). In cases where examples are more than one word (e.g., that clauses controlled by a verb), the entire "that" clause is bolded, and the controlling word is underlined.

Table 17 *To* Clauses

Feature	Description/example
To clauses (all)	*You can save your reflection* ***to remind yourself*** *and stay motivated throughout the course*
To clauses controlled by a verb	*There are many such configurations that have been* ***studied to address the problem of open configurations***
Speech act verb	*claim*, *report*, *show*, *warn*
Cognition verb	*estimate*, *imagine*, *presume*, *suppose*
Desire/intent/decision verb	*consent*, *intend*, *prefer*, *refuse*
Modality/cause/effort verb	*attempt*, *counsel*, *defy*, *persuade*
Probability/simple fact verb	*appear*, *happen*, *seem*, *tend*
To clauses controlled by an adjective	*Vocabulary is* ***easier to acquire*** *when there are contextual clues to help convey meaning*
Certainty adjectives	*apt*, *due*, *prone*, *sure*
Ability/willingness adjectives	*able*, *disposed*, *hesitant*, *willing*
Personal affect adjectives	*astonished*, *disgusted*, *pleased*, *relieved*
Ease/difficulty adjectives	*easier*, *hard*, *possible*, *tough*
Evaluative adjectives	*convenient*, *desirable*, *inappropriate*, *useful*
To clauses controlled by a noun	*I would think making complex characters requires the* ***ability to place oneself into the figurative shoes of another person, into their mind and body***

Note. Bold indicates the target item (e.g., adjective). In cases where examples are more than one word (e.g., that clauses controlled by a verb), the entire “that” clause is bolded, and the controlling word is underlined.

Table 18 Summary of Statistical Details of Previous Multidimensional Analyses

EFA methodological choices	Biber (1988)	Biber et al. (2004) and Biber (2006)	Staples et al. (2017)	Yan and Staples (2020)
Factor extraction	Principal axis	Principal axis	Principal axis	Principal axis
Criteria for determining the number of dimensions	Scree plot; variance explained	Scree plot; variance explained	Scree plot; variance explained	Scree plot; variance explained
Factor rotation	Promax	Promax	Promax	Promax
Cutoff loading	|.35|	|.30|	|.30|	|.30|
Cutoff communality	None	.15	None	None

we use *dimension* to cover the substantive interpretation of each constellation of linguistic features in light of the register differences (e.g., oral vs. literate discourse; see Biber et al., 2004).

Exploratory Factor Analysis

An EFA inherently entails a series of methodological decisions that ultimately affect overall results (Brown, 2015; Fabrigar et al., 1999; Loewen & Gonulal, 2015; Sawaki, 2013). These include, but are not limited to, the selection of (a) factor extraction method, (b) the criteria for deciding the number of factors, (c) the factor rotation method, and (d) the cutoff loading and communality values used to retain variables. Table 18 summarizes the methodological choices made in a selection of previous MDAs to closely replicate the procedures. For implementation, *R* equivalents for each analysis step were identified from the *psych* package (Revelle, 2016).

For the factor extraction method, principal axis factoring (PA) was selected and implemented with “PA” in the *fa*( ) function in the *psych* package (Revelle, 2016). PA is considered to be relatively robust to the violation of the multivariate normality assumption (Brown, 2015; Fabrigar & Wegener, 2012) and thus was deemed appropriate given the nature of corpus data. The decision on the appropriate number of factors was guided by multiple criteria: a scree plot, the total amount of variance explained, and the substantive interpretation of the specific factor solution (see Brown, 2015; Fabrigar & Wegener, 2012; Loewen & Gonulal, 2015; Sawaki, 2013). Once possible factor solutions were identified, a Promax rotation was used, which allowed the latent dimensions of linguistic features to be correlated (Fabrigar & Wegener, 2012;

Sawaki, 2013). Following previous studies (e.g., Biber, 2006; Biber et al., 2004), the rotated pattern matrices, the amount of explained variance, and the result of scree plot were used to decide the final solution in our study.

Dimension Scores

The second step in MDA was to calculate dimension scores. Following previous MDAs, dimension scores were calculated using the linguistic features that satisfied the cutoff loading and communality criteria. In this study, cutoff loading of |.30| and communality of .15 were used as the criteria for retaining linguistic features (Biber, 2006, p. 183). While a communality criterion has not always been used in prior MDAs (see Table 18), the use of such a criterion helps ensure that the variance of each linguistic feature in the MDAs is well explained by the latent factors (Brown, 2015; Fabrigar & Wegener, 2012). Following previous MDAs, linguistic features with cross loading were counted only once in the primary factor (Biber, 1988, 2006; Biber et al., 2004). Subsequent to this filtering process, linguistic features were each standardized across the texts in the entire corpus using the Z-scores. When it comes to understanding the relative standings of each document in our corpus with regard to the identified linguistic dimensions, using absolute frequencies to calculate dimensional scores will overestimate the contributions of highly frequent linguistic features (for an elaborated explanation of this rationale, see Biber, 1988). Once standardized, for each factor, the sum of negatively loaded features (in the standardized scale) was subtracted from the sum of positively loaded features (in the standardized scale). Accordingly, the dimensional score of a text on a dimension is considered to be the relative distance (in a standardized scale) from the mean of the entire corpus on that particular dimension (for examples of this process, see Biber, 1988, 2006).

Comparison of Spoken and Written Language Between the Technology-Mediated Learning Environment and TOEFL 2000 Spoken and Written Academic Language Corpora

The dimension scores for each text were then used to make a series of comparisons with regard to spoken and written language use within and across learning environments (TMLE and T2K-SWAL). To make the comparisons between spoken TMLE texts, written TMLE texts, spoken T2K-SWAL texts, and written T2K-SWAL texts, we created a multilevel regression model (Gelman & Hill, 2007; Heck & Thomas, 2020; Hox, 2018).

The multilevel linear models were constructed using the *lme4* package (Bates et al., 2015) in *R* (R Development Core Team, 2014), as follows. Because the *lme4* package does not allow a multivariate outcome variable, we included dimension as the first predictor in the baseline model we constructed. Also included in this baseline model was learning environment (TMLE = 0; traditional = 1). Thus our baseline model simply compared the differences between TMLE and T2K-SWAL in terms of the dimension scores. To determine whether differences existed with regard to linguistic features across learning environment and/or mode, we added the main effect of mode (spoken = 0; written = 1) and three-way interactions of dimension, mode, and learning environment. To control for the confounding effects of other situational variables, we have included the random intercepts of each discipline-by-mode combination and slopes of dimension and learning environment for these grouping variables and by-text-type random intercepts and slopes of dimension. This procedure was implemented to guard against Type 1 error(s) and to account for biased fixed-effects estimates due to unbalanced numbers of texts from our subcorpora (Barr et al., 2013; Hox, 2018). Effect sizes (marginal and conditional R^2 values) were calculated using the *MuMIn* package (Bartoń, 2019). The *R* code for our model was as follows:

```
lmer(Dimensional_score ~ Dimension * learning_environment * mode+
 (0+dummy(Dimension, “DIM1”)|filename) +
(1+dummy(Dimension,“DIM1”)+dummy(learning_environment,“tmle”)|discipline:mode)+
 (1+dummy(Dimension,“DIM1”)|text_type), REML = F)
```

When models did not converge, different optimizers were tested to achieve model convergence. If the model still did not converge, the complexity of the random structure was reduced (Barr et al., 2013). The final regression model was submitted to post hoc comparisons (Tukey’s HSD) through the *esmeans* package (Lenth et al., 2018), which allowed us to obtain model-based marginal means, confidence intervals, and effect sizes. For ease of interpretation, results were also visualized through the *emmeans::plot*( ) function. Effect sizes (Cohen’s d) for each comparison were estimated through *emmeans::eff_size*( ) function in the same package. For interpretation of effect sizes, we followed Plonsky and Oswald’s (2014) benchmarks (small = .40; medium = .70; large = 1.00).

To paint a detailed picture of the distributional differences, we have also plotted the raw data for each comparison (Figures 3–7). These plots combine box plot, violin plot, and jitters to summarize the distributional details in a digestible format. The jitter shows the individual data points, the outer curves (i.e., violin) show the distributional density of the data, and the box plot summarizes the central tendencies in a concise manner.

Results

Multidimensional Analysis

Initially, 133 linguistic features were considered for inclusion in the EFA. Preliminary analyses indicated that 47 indices were particularly rare in the corpus (i.e., had zero counts for a large proportion of texts), and these were removed from further analysis following previous studies (Biber, 2006; Biber et al., 2004). When possible, these linguistic features were combined into larger categories (e.g., phrasal verbs). In other cases, the linguistic feature had to be removed from the EFA to mitigate the adverse effects of zero-inflated distributions. After the preliminary filtering process, 86 of the original 133 linguistic features were included in the EFA.

The data set was shown to be adequate for an EFA (Kaiser–Meyer–Olkin measure of sampling adequacy [KMO] = .816; Bartlett's $K^2 = 1{,}470{,}539$), $p < .001$. A visual inspection of a scree plot suggested that the appropriate number of factors was between five and seven (see Figure 1), which did not differ greatly in the amount of variance explained (29%, 31%, and 33%, respectively). To understand the nature of each solution, the pattern matrices using Promax rotation were compared for the substantive interpretation, including the four-factor solution corresponding to the Biber's (2006; see also Biber et al., 2004) factor solution on the T2K-SWAL corpus. Given this information, we concluded that the six-factor solution yielded the most straightforward interpretation of linguistic dimensions while explaining an acceptable amount of variance in the entire data set (31%). The interfactor correlations are presented in Table 19. In what follows, we describe each dimension of the linguistic features based on the six factors.

Dimension 1: Oral Versus Literate Discourse

The first dimension, much like previous MDAs (e.g., Biber, 1988; Biber et al., 2004), distinguishes between features of oral discourse (positive features) and features of literate discourse (negative features). The strongest positive features include the use of clausal coordination, contractions, emphatics, copular *be* constructions, demonstrative pronouns, and first person pronouns, all of which are characteristic of (relatively) unplanned spoken language. The negative features include a higher proportion of nonfinite clauses, use of nouns, more complex clauses and T-units, and more varied lexical use. It should be noted that the current analysis included fewer features for Dimension 1 than the original T2K-SWAL MDA (Biber et al., 2004), although substantial overlap exists, particularly for positively loaded features with stronger loadings. Although some overlap also exists with the negative features, most of the negatively loaded features from the T2K-SWAL MDA are in Dimension 2 of the current analysis. See Tables 20–25.

Dimension 2: Lexical and Phrasal Complexity

Dimension 2 includes features that are primarily related to lexical and phrasal complexity—features that have been shown to be representative of academic writing (e.g., Biber et al., 2011). The positive features for this dimension include the number of characters per word (which is a proxy for lexical sophistication); phrasal coordination; and the use of attributive adjectives, nominalizations, and prepositional phrases. Only one feature, conditional adverbial subordinators (*if*, *unless*), loaded negatively on this dimension.

Dimension 3: Procedural Discourse

Dimension 3 is related to procedural language and shares a number of features with Biber et al.'s (2004) second dimension. Positive features include the use of infinitive clauses, a higher proportion of verb use, second person pronouns, finite dependent clauses, and the use of possessives. This dimension has no negative features.

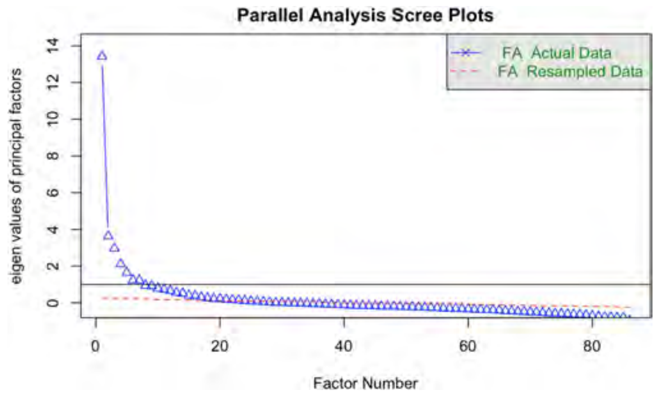

Figure 1 Parallel analysis scree plots.

Table 19 Interfactor Correlations for the Final Six-Factor Solution

Dimension	Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6
Dimension 1	1					
Dimension 2	−.217	1				
Dimension 3	.382	−.214	1			
Dimension 4	.398	−.042	.106	1		
Dimension 5	.210	.071	.136	.128	1	
Dimension 6	.134	.274	−.081	.289	−.07	1

Dimension 4: Elaborated Discourse—Complement Clauses

Dimension 4 highlights the use of clausal complements, in particular, *that* clausal complements controlled by a verb. This dimension has no negative features.

Dimension 5: Narrative Orientation

The linguistic features of this dimension overlap with Biber et al.'s (2004) third dimension and highlight features related to narration, including the use of third person pronouns, past tense verbs, and animate nouns. This dimension has no negative features.

Dimension 6: Elaborated Discourse—Relative Clauses

This dimension highlights the use of relative clauses (in particular, WH-relative clauses). This dimension has no negative features.

Patterns of Variation Across Learning Environments, Modes, and Registers

After identifying the six linguistic dimensions of language use in the combined TMLE/T2K-SWAL corpus, we investigated the extent to which features of spoken and written language in technology-mediated and traditional learning environments were (dis)similar with regard to these linguistic dimensions. A series of multilevel models revealed that the fixed effects of dimension, learning environment, and mode and their three-way interaction significantly improved the model fit compared to the baseline model, $\chi^2(12) = 2{,}867.6$, $p < .001$ (ΔBIC = −2,745.014; ΔAIC = −2,843.624). The random effect structure proposed in the Method section converged with the default optimizer. The marginal (fixed effects only) and conditional (random effects included) R^2 were .393 and .583, respectively.

In what follows, we present pairwise comparisons of model-based marginal means with regard to scores on each dimension. In each section, we report and visualize the descriptive statistics, visualize a summary of the pairwise comparisons,

Table 20 Factor Loadings for Dimension 1

Linguistic feature	Loading	Communality
Positive features		
Independent clause coordination	.774	.587
Contractions	.754	.641
Emphatics	.750	.543
Be as main verb	.732	.549
Demonstrative pronouns	.634	.557
First person pronouns	.613	.481
It pronouns	.585	.426
Factive adverbs	.557	.307
That relative clauses	.348	.328
Determiners per nominal	.341	.356
Causative adverbial subordinators	.308	.163
Negative features		
Nonfinite clauses	−.813	.725
Nouns	−.619	.816
Mean verbal dependencies	−.568	.370
Mean length of clause	−.484	.539
Abstract nouns	−.442	.366
Mean length of T-unit	−.433	.331
Moving-average type–token ratio	−.411	.299

Table 21 Factor Loadings for Dimension 2

Linguistic feature	Loading	Communality
Positive features		
Word length	.588	.798
Phrasal coordination	.561	.414
Coordinating conjunctions per nominal	.526	.281
Nominalizations	.498	.466
Attributive adjectives	.441	.406
Adjectival modifiers	.373	.194
Prepositional phrases	.354	.367
Prepositions per nominal	.301	.245
Negative feature		
Conditional adverbial subordinator	−.330	.179

Table 22 Factor Loadings for Dimension 3

Linguistic feature (positive only)	Loading	Communality
To clauses	.803	.555
Verbs	.695	.795
Non–past tense	.665	.797
To clause verbs	.515	.399
Second person pronouns	.514	.354
Mental verbs	.505	.385
To clause with verbs of desire	.447	.241
To clause nouns	.439	.212
Dependent clause per clauses	.413	.508
Activity verbs	.391	.296
Possessives	.372	.274

Table 23 Factor Loadings for Dimension 4

Linguistic feature (positive only)	Loading	Communality
That complement clauses controlled by a verb	.886	.737
That complement clauses	.839	.764
That complement clauses controlled by a factive verb	.787	.546
All clausal complements	.469	.396

Table 24 Factor Loadings for Dimension 5

Linguistic feature (positive only)	Loading	Communality
Third person pronouns	.611	.431
Past tense verbs	.577	.404
Animate nouns	.439	.288

Table 25 Factor Loadings for Dimension 6

Linguistic feature (positive only)	Loading	Communality
WH-relative clauses	.824	.710
Relative clauses per clause	.807	.696
WH-relative subject clauses	.708	.563
Relative clauses per nominal	.596	.715

and report on selected pairwise comparisons that highlight differences (or lack thereof) between spoken and written language across the learning environments.

Dimension 1: Oral Versus Literate Discourse

As noted in the previous sections, Dimension 1 contrasts oral discourse with literate discourse. Accordingly, spoken texts across learning environments tend to earn positive scores on this dimension, while written texts tend to earn negative scores. See Table 26 for descriptive statistics, and see Figure 2 for a visualization of these data.

Pairwise comparisons indicate significant and meaningful differences across modes and learning environments with regard to Dimension 1 scores (see Table 27 for a summary of the results). Spoken texts earned significantly higher Dimension 1 scores (with large effect sizes) than written texts in both TMLEs, $p < .001$, $d = 4.002$, and traditional learning environments, $p < .001$, $d = 6.033$. Written texts did not differ across learning environments, $p = .580$, $d = -.503$, but TMLE spoken texts earned significantly lower Dimension 1 scores (with a large effect) than T2K-SWAL spoken texts, $p < .001$, $d = -2.534$. These results indicate that, on average, spoken TMLE texts include fewer features traditionally related to oral discourse. As indicated in Figure 2, some TMLE texts earned similar Dimension 1 scores (i.e., had a similar proportion of linguistic features related to oral discourse), but a large proportion of TMLE texts earned lower scores. This result may indicate that, from a strictly linguistic standpoint, many TMLE texts may be more difficult to comprehend than spoken texts in traditional environments (for a discussion on the relationship between linguistic features and listening difficulty, see, e.g., Brunfaut & Révész, 2015).

Dimension 2: Lexical and Phrasal Complexity

Dimension 2 scores highlight lexical and phrasal complexity. As found in previous research, spoken texts across learning environments tend to earn negative scores for this dimension, while written texts tend to earn positive scores for this dimension. See Table 28 for descriptive statistics and Figure 3 for a visualization of these data.

The pairwise comparisons (see Table 29 for a summary of the results) indicate that spoken texts earn significantly lower Dimension 2 scores than written texts in both TMLEs, $p < .001$, and traditional learning environments, $p < .001$,

Table 26 Dimension 1 Scores by Mode and Learning Environment

		Descriptive statistics			Model-based marginal means	
Learning environment	Mode	*n*	*M*	*SD*	*M*	*SE*
TMLE	Spoken	2, 233	6.659	7.601	5.902	1.841
T2K-SWAL	Spoken	290	14.291	4.531	16.201	1.249
TMLE	Written	1,868	−9.446	7.821	−10.361	1.154
T2K-SWAL	Written	172	−7.954	5.300	−8.316	1.423

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

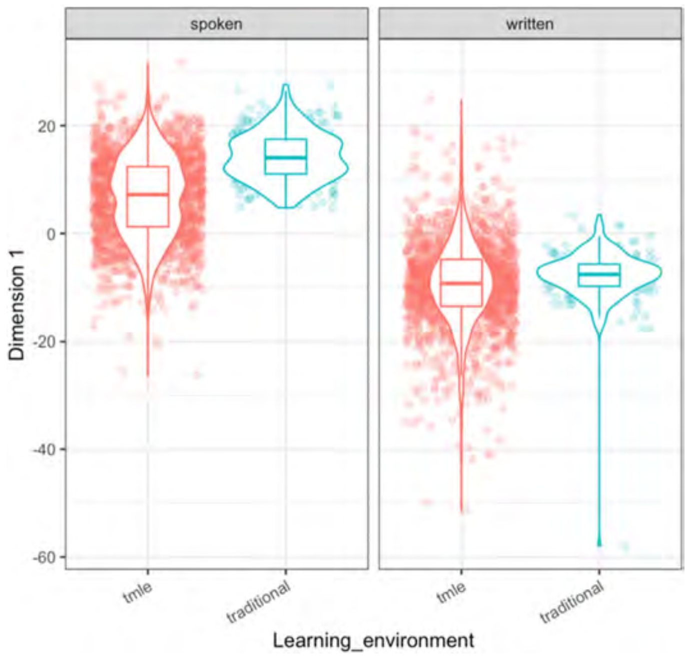

Figure 2 Distribution of Dimension 1 scores by mode and learning environment.

Table 27 Pairwise Comparisons of Estimated Marginal Means for Modes (Dimension 1)

Contrast (Item 1/Item 2)	Estimate	*SE*	*Z*-ratio	*p*	*d*	*SE* of *d*
TMLE – spoken/T2K-SWAL – spoken	−10.299	2.037	−5.056	<.001	−2.534	.501
TMLE – spoken/TMLE – written	16.263	2.173	7.484	<.001	4.002	.535
TMLE – spoken/T2K-SWAL – written	14.218	2.327	6.111	<.001	3.499	.573
T2K-SWAL – spoken/TMLE – written	26.562	1.701	15.617	<.001	6.536	.419
T2K-SWAL – spoken/T2K-SWAL – written	24.517	1.893	12.950	<.001	6.033	.466
TMLE – written/T2K-SWAL – written	−2.045	1.606	−1.273	.580	−.503	.395

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

though the difference is larger in traditional learning environments, $d = -3.049$, than in TMLEs, $d = -.691$. The pairwise comparisons also indicate that spoken and written texts differ significantly across learning environments. Spoken TMLE texts earn significantly higher Dimension 2 scores than spoken T2K-SWAL texts, $p = .001$, $d = 1.228$, indicating that they include more complex lexical and phrasal features. Conversely, written TMLE texts earn significantly lower Dimension 2 scores than written T2K-SWAL texts, $p = .001$, $d = -1.131$, indicating that written T2K-SWAL texts include more complex lexical and phrasal features. A preliminary explanation for these findings may be that spoken TMLE texts,

Table 28 Descriptive Statistics for Spoken and Written Texts (Dimension 2)

		Descriptive statistics			Model-based marginal means	
Learning environment	Mode	*n*	*M*	*SD*	*M*	*SE*
TMLE	Spoken	2,233	−1.170	4.658	−1.245	.477
T2K-SWAL	Spoken	290	−6.048	2.991	−6.234	.402
TMLE	Written	1,868	1.766	5.551	1.562	.338
T2K-SWAL	Written	172	6.207	3.935	6.159	.468

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

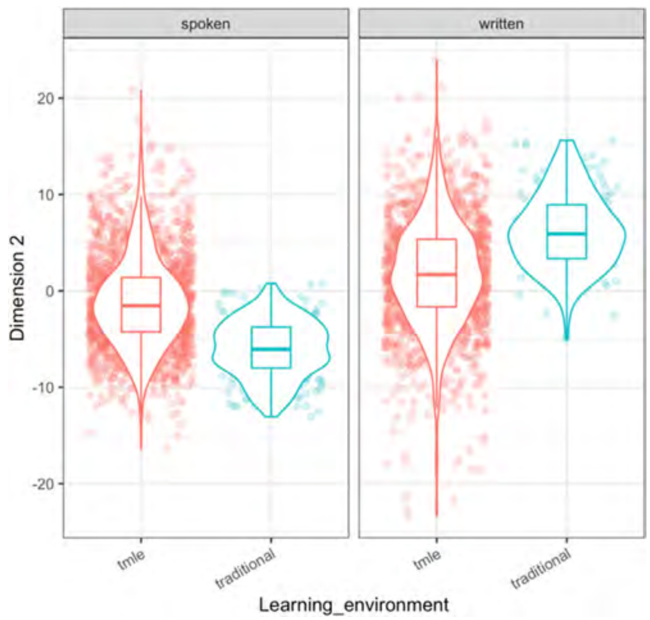

Figure 3 Distribution of Dimension 2 scores by mode and learning environment.

Table 29 Pairwise Comparisons of Estimated Marginal Means for Modes (Dimension 2)

Contrast (Item 1/Item 2)	Estimate	*SE*	*Z*-ratio	*p*	*d*	*SE* of *d*
TMLE – spoken/T2K-SWAL – spoken	4.989	.535	9.321	<.001	1.228	.132
TMLE – spoken/TMLE – written	−2.807	.585	−4.800	<.001	−.691	.144
TMLE – spoken/T2K-SWAL – written	−7.404	.668	−11.082	<.001	−1.822	.164
T2K-SWAL – spoken/TMLE – written	−7.796	.525	−14.843	<.001	−1.918	.129
T2K-SWAL – spoken/T2K-SWAL – written	−12.392	.617	−20.100	<.001	−3.049	.152
TMLE – written/T2K-SWAL – written	−4.597	.482	−9.530	<.001	−1.131	.119

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

which comprise recorded lectures and other videos, likely involve more planning than the spoken T2K-SWAL texts (e.g., in-person lectures, office hours, and service encounters). Conversely, many written online sources are created in a less formal environment (with less editing and revising) than many of the written T2K-SWAL texts (e.g., textbooks), resulting in less complex language. From a practical standpoint, these results may suggest that from a linguistic perspective, TMLE spoken texts may (on average) be more difficult to comprehend than traditional spoken texts and that the opposite may be true for written texts across the two environments (see, e.g., Crossley et al., 2008).

Table 30 Descriptive Statistics for Spoken and Written Texts (Dimension 3)

		Descriptive statistics			Model-based marginal means	
Learning environment	Mode	*n*	*M*	*SD*	*M*	*SE*
TMLE	Spoken	2,233	1.823	5.306	1.748	.477
T2K-SWAL	Spoken	290	2.531	3.303	2.346	.402
TMLE	Written	1,868	−2.132	7.416	−2.335	.338
T2K-SWAL	Written	172	−4.775	3.764	− 4.824	.468

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

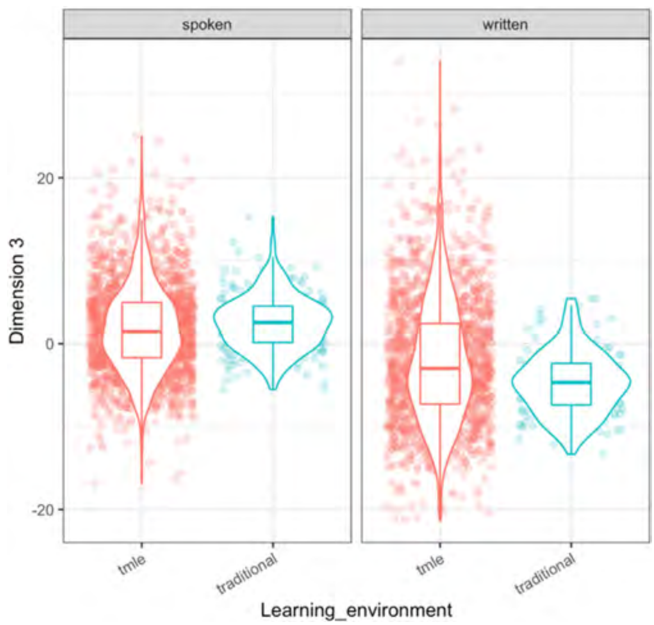

Figure 4 Distribution of Dimension 3 scores by mode and learning environment.

Dimension 3: Procedural Discourse

Dimension 3 scores highlight language related to procedural discourse. Spoken texts tend to earn positive scores on this dimension, while written texts tend to earn negative scores. Descriptive statistics for Dimension 3 are included in Table 30 and visualized in Figure 4.

The pairwise comparisons (see Table 31 for a summary of results) indicate that spoken texts earn significantly higher Dimension 3 scores than written texts in both TMLEs, $p < .001$, $d = 1.005$, and traditional learning environments, $p < .001$, $d = 1.764$. Spoken TMLE texts earned slightly lower scores than spoken T2K-SWAL texts, but these differences were not significant, $p = .678$, $d = -.147$. Conversely, written TMLE texts earned significantly higher scores than written T2K-SWAL texts, $p < .001$, $d = .612$. These results indicate that, on average, written TMLE texts include more procedural discourse features (such as those included in tutorials and how-to articles) than written texts in traditional environments.

Dimension 4: Elaborated Discourse—Clausal Complements

Dimension 4 scores highlight discourse that is elaborated through the use of clausal complements. Spoken texts on this dimension tend to earn positive scores, while written texts tend to earn negative scores. Descriptive statistics for Dimension 4 scores are included in Table 32 and visualized in Figure 5.

Table 31 Pairwise Comparisons of Estimated Marginal Means for Modes (Dimension 3)

Contrast (Item 1/Item 2)	Estimate	*SE*	*Z*-ratio	*p*	*d*	*SE* of *d*
TMLE – spoken/T2K-SWAL – spoken	−.599	.535	−1.118	.678	−.147	.132
TMLE – spoken/TMLE – written	4.083	.585	6.981	<.001	1.005	.144
TMLE – spoken/T2K-SWAL – written	6.571	.668	9.836	<.001	1.617	.164
T2K-SWAL – spoken/TMLE – written	4.682	.525	8.914	<.001	1.152	.129
T2K-SWAL – spoken/T2K-SWAL – written	7.170	.617	11.629	<.001	1.764	.152
TMLE – written/T2K-SWAL – written	2.488	.482	5.159	<.001	.612	.119

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

Table 32 Descriptive Statistics for Spoken and Written Texts (Dimension 4)

		Descriptive statistics			Model-based marginal means	
Learning environment	Mode	*n*	*M*	*SD*	*M*	*SE*
TMLE	Spoken	2,233	1.206	3.447	1.131	.477
T2K-SWAL	Spoken	290	.074	1.311	−.112	.402
TMLE	Written	1,868	−1.352	3.009	−1.555	.338
T2K-SWAL	Written	172	−1.108	1.736	−1.157	.468

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

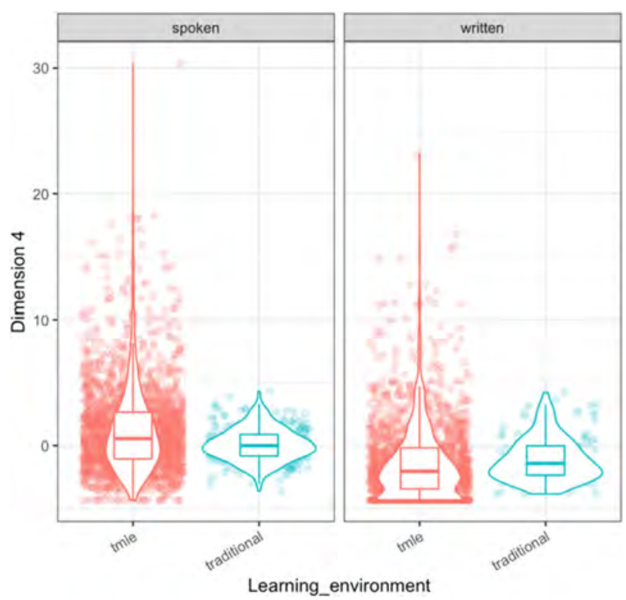

Figure 5 Distribution of Dimension 4 scores by mode and learning environment.

The pairwise comparisons (see Table 33 for a summary of results) indicate that spoken texts earn significantly higher Dimension 4 scores (with a small effect) than written texts in TMLEs, $p < .001$, $d = .661$, but not in traditional learning environments, $p = .326$, $d = .257$. No significant differences were found between learning environments for spoken, $p = .093$, $d = .306$, or written texts, $p = .843$, $d = -.098$.

Table 33 Pairwise Comparisons of Estimated Marginal Means for Modes (Dimension 4)

Contrast (Item 1/Item 2)	Estimate	*SE*	*Z*-ratio	*p*	*d*	*SE* of *d*
TMLE – spoken/T2K-SWAL – spoken	1.243	.535	2.322	.093	.306	.132
TMLE – spoken/TMLE – written	2.686	.585	4.593	<.001	.661	.144
TMLE – spoken/T2K-SWAL – written	2.288	.668	3.425	.003	.563	.164
T2K-SWAL – spoken/TMLE – written	1.443	.525	2.748	.031	.355	.129
T2K-SWAL – spoken/T2K-SWAL – written	1.046	.617	1.696	.326	.257	.152
TMLE – written/T2K-SWAL – written	−.398	.482	−.825	.843	−.098	.119

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

Table 34 Descriptive Statistics for Spoken and Written Texts (Dimension 5)

		Descriptive statistics			Model-based marginal means	
		n	*M*	*SD*	*M*	*SE*
TMLE	Spoken	2,233	.304	2.242	.229	.477
T2K-SWAL	Spoken	290	.668	1.753	.483	.402
TMLE	Written	1,868	−.518	2.142	−.722	.338
T2K-SWAL	Written	172	.559	2.016	.511	.468

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

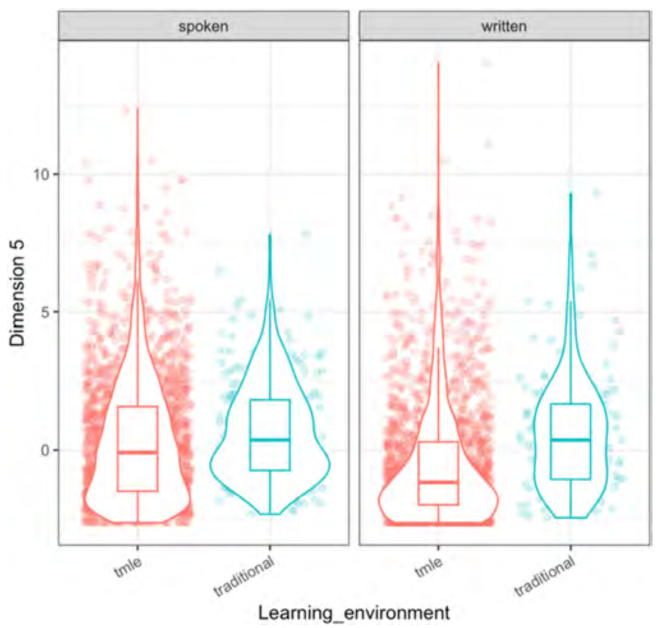

Figure 6 Distribution of Dimension 5 scores by mode and learning environment.

Dimension 5: Narrative Orientation

Dimension 5 scores highlight features related to narration (i.e., third person pronouns, past tense verbs, and animate nouns). Descriptive statistics for Dimension 5 scores are included in Table 34 and visualized in Figure 6.

Pairwise comparisons indicate that no significant differences were found across spoken and written texts in either TMLEs, $p = .364$, $d = 0.234$, or traditional learning environments, $p = 1.000$, $d = -0.007$. These comparisons also indicated that there were no significant differences across learning environments for either spoken, $p = .965$, $d = -0.063$, or written texts, $p = .052$, $d = -0.303$. See Table 35.

Table 35 Pairwise Comparisons of Estimated Marginal Means for Modes (Dimension 5)

Contrast (Item 1/Item 2)	Estimate	*SE*	*Z*-ratio	*p*	*d*	*SE* of *d*
TMLE – spoken/T2K-SWAL – spoken	−.254	.535	−.475	.965	−.063	.132
TMLE – spoken/TMLE – written	.951	.585	1.625	.364	.234	.144
TMLE – spoken/T2K-SWAL – written	−.282	.668	−.422	.975	−.069	.164
T2K-SWAL – spoken/TMLE – written	1.205	.525	2.294	.099	.297	.129
T2K-SWAL – spoken/T2K-SWAL – written	−.028	.617	−.045	1.000	−.007	.152
TMLE – written/T2K-SWAL – written	−1.233	.482	−2.555	.052	−.303	.119

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

Table 36 Descriptive Statistics for Spoken and Written Texts (Dimension 6)

		Descriptive statistics			Model-based marginal means	
Learning environment	Mode	*n*	*M*	*SD*	*M*	*SE*
TMLE	Spoken	2,233	.846	3.001	.771	.477
T2K-SWAL	Spoken	290	−.491	1.558	−.677	.402
TMLE	Written	1,868	−.932	3.279	−1.136	.338
T2K-SWAL	Written	172	−.022	1.584	−.070	.468

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

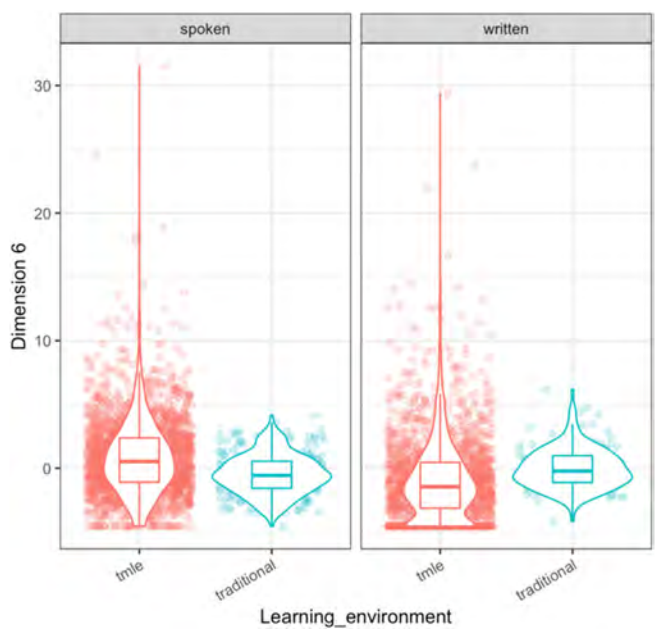

Figure 7 Distribution of Dimension 6 scores by mode and learning environment.

Dimension 6: Elaborated Discourse—Relative Clauses

Dimension 6 highlights the use of relative clauses. Descriptive statistics for Dimension 6 scores are included in Table 36 and visualized in Figure 7.

Pairwise comparisons (see Table 37 for a summary of these results) indicate that spoken texts earn significantly higher Dimension 6 scores than written texts (with a small effect) in TMLEs, $p = .006$, $d = 0.469$, but not in traditional learning environments, $p = .759$, $d = -0.149$. Spoken TMLE texts also earned significantly higher Dimension 6 scores than spoken

Table 37 Pairwise Comparisons of Estimated Marginal Means for Modes (Dimension 6)

Contrast (Item 1/Item 2)	Estimate	*SE*	*Z*-ratio	*p*	*d*	*SE* of *d*
TMLE – spoken/T2K-SWAL – spoken	1.447	.535	2.704	.035	.356	.132
TMLE – spoken/TMLE – written	1.906	.585	3.259	.006	.469	.144
TMLE – spoken/T2K-SWAL – written	.841	.668	1.258	.590	.207	.164
T2K-SWAL – spoken/TMLE – written	.459	.525	.874	.818	.113	.129
T2K-SWAL – spoken/T2K-SWAL – written	−.607	.617	−.984	.759	−.149	.152
TMLE – written/T2K-SWAL – written	−1.066	.482	−2.209	.121	−.262	.119

Note. T2K-SWAL = TOEFL 2000 Spoken and Written Academic Language corpus; TMLE = Technology-Mediated Learning Environment corpus.

T2K-SWAL texts, $p = .035$, $d = 0.356$, but with a negligible effect size. No significant differences were found across learning environments for written texts.

Discussion

This study reports on an effort to collect a corpus of texts used in TMLEs, describes the linguistic features of those texts, and determines the degree to which these texts differed from texts used in traditional learning environments (represented by the T2K-SWAL corpus). A summary of the results is included herein, followed by implications for the TOEFL iBT test, potential limitations, and future directions.

Summary of Results

An EFA was used to identify six latent factors based on linguistic features related to lexical complexity, grammatical complexity, and features of lexicogrammatical use. The first dimension (oral vs. literate discourse) clearly discriminated between spoken and written modes but also indicated that TMLE spoken registers (course management talk and instructional videos) earned significantly lower scores than T2K-SWAL spoken registers. This finding suggests that spoken TMLE texts may be more difficult to comprehend than spoken texts in traditional learning environments. Dimension 2 (lexical and phrasal complexity) discriminated between spoken and written modes within each learning environment and also discriminated between spoken and written modes across learning environments. The results also indicate that spoken TMLE texts may be more difficult to comprehend than spoken texts in traditional learning environments. Furthermore, the results indicate that TMLE written texts may be easier to comprehend than written texts in traditional learning environments. Dimension 3 (procedural discourse) highlighted differences between spoken and written texts across learning environments (wherein spoken texts tend to include more procedural language). Additionally, Dimension 3 highlighted differences in written texts across learning environments—written TMLE texts included significantly more linguistic features related to procedural discourse than written texts in traditional learning environments.

Dimensions 4, 5, and 6 were less useful in discriminating between modes and learning environments. Dimension 4 (elaborated discourse—complement clauses) highlighted minor differences between spoken and written TMLE texts (spoken texts included more complement clauses), but no significant differences were found across learning environments. No significant differences were found with regard to Dimension 5 (narrative orientation) scores across modes or learning environments. With regard to Dimension 6 (elaborated discourse—relative clauses), minor significant differences were found across modes in TMLEs (spoken texts included more relative clauses) but not in traditional learning environments. Spoken TMLE texts were also found to have significantly higher Dimension 6 scores than spoken traditional texts, but the effect size was negligible.

Implications

Given the growing use of TMLEs in higher education, both in online and hybrid courses and in “normal” courses, it is important to outline the linguistic features of texts encountered in these environments and how (and the degree to which) they differ from texts that have traditionally been encountered in university courses (which are represented by T2K-SWAL). The primary purpose of this study was to determine the degree to which TMLE texts differed from texts in traditional learning environments (represented by T2K-SWAL) across spoken and written modes to preliminarily determine whether assessment tasks with new features (e.g., with more varied text characteristics) may be needed to bolster validity arguments for the TOEFL iBT test (e.g., with regard to the domain description and extrapolation inferences). The results indicate that there are indeed some differences between spoken and written language use across learning environments. For example, TMLE spoken texts (e.g., instructional and course management videos) have fewer features related to oral discourse (e.g., coordinated clauses, contractions, and emphatics) than spoken texts in traditional learning environments. Previous research (e.g., Brunfaut & Révész, 2015) has indicated that spoken texts with a higher proportion of oral discourse features (e.g., the use of contractions) are easier to comprehend. The findings of this study, in light of previous research on listening comprehension, suggest that spoken texts in TMLEs may be more difficult to comprehend than those in traditional learning environments. Furthermore, the results indicate that spoken TMLE texts tended to include more and more complex lexical and phrasal features than T2K-SWAL spoken texts (e.g., lectures, office hours, and service encounters), which may also indicate that TMLE spoken texts may be more difficult to comprehend than spoken texts in traditional learning environments (e.g., Crossley et al., 2008). Preliminarily, the findings of this study in light of previous research on reading difficulty suggest that assessment tasks designed to align with the features of spoken academic language in traditional learning environments may be easier to complete successfully than some listening tasks in TMLEs, which may affect support for the extrapolation inference. However, more research is needed to determine (a) the specific spoken TMLE registers that are contributing to the lower mean scores for Dimensions 1 and 2, (b) the prevalence of these registers in the university experience, and (c) the degree to which the inclusion/absence of these features affects task difficulty.

The results also indicate that written TMLE texts tended to be less complex with regard to lexical (e.g., word length, use of nominalizations) and phrasal (e.g., coordinated noun phrases) features than written texts in traditional learning environments. Previous reading difficulty research (e.g., Crossley et al., 2008; Kincaid et al., 1975) has indicated that these features contribute to reading difficulty, suggesting that written TMLE texts are (on average) easier to comprehend than texts in traditional learning environments. Preliminarily, this may suggest that assessment tasks designed to align with the features of written traditional learning environments may be more difficult to complete than some TMLE tasks. However, as with the preliminary results regarding spoken texts, more research is needed to determine (a) the specific written registers that have lower Dimension 2 scores, (b) the prevalence of these texts in the university experience, and (c) the degree to which the inclusion/exclusion of these features affects task difficulty.

Limitations and Future Directions

This study described the collection of a new corpus of academic language from TMLEs and a preliminary linguistic analysis of the similarities and differences between spoken and written language use within and across learning environments using a new MDA. Future research should build on this study by addressing the following limitations. First, this study focused on differences and similarities across modes but did not explicitly investigate the degree to which language varied across specific spoken and written registers. Given the wide distribution of dimension scores (particularly across TMLE texts) and that research has indicated that language use can also vary by specific registers, future research should examine the characteristics of specific registers. Second (and relatedly), we did not examine differences across disciplines. Future research should investigate these differences and their implications.

In this study, every effort was made to ensure that the texts collected were representative of the texts used in TMLEs. However, in the end, fairly low participation across institutions resulted in a convenience sample (which is common in related studies). Additionally, because of relatively low participation at the selected institutions, the corpus was supplemented with texts from MOOCs. While the use of MOOC data allowed for greater control over the representativeness of the texts collected (we sampled from a wide variety of courses and registers), texts used in MOOCs may differ to some degree from those used in other online, hybrid, or technology-enhanced courses. Future research should investigate the degree to which MOOCs are representative of texts used in other TMLEs. Additionally, the TMLE corpus described in this manuscript and T2K-SWAL represent texts that are encountered by students in university learning environments and are useful for modeling the linguistic characteristics of TOEFL iBT test tasks that assess receptive skills. They may not, however, represent the types of texts that students *produce* in university learning environments, and therefore other corpora may be needed to model this type of language (particularly in TMLEs). With regard to the statistical procedure in conducting the MDA, the present study conducted an EFA followed by a manual calculation of dimensional scores to enhance the consistency with previous MDAs (e.g., Biber, 1988). However, as one of the reviewers recommended, it is also possible to calculate dimensional scores directly from model estimates and factor loadings through, for example, confirmatory factor analysis. Thus future studies may benefit from comparing these two approaches of deriving dimension scores.

Conclusion

This study described the collection of a new corpus of academic language encountered in university-level TMLEs that is designed to supplement the previously collected T2K-SWAL (Biber et al., 2004), which represents spoken and written language encountered in traditional learning environments. The linguistic features of TMLEs were outlined using a MDA (Biber, 1988), and features of spoken and written language use in TMLEs were compared and contrasted with features of spoken and written language use in T2K-SWAL. The results indicate that although a number of similarities exist across the two learning environments, large, significant differences are also evident, particularly with regard to spoken language. These differences may have important implications for the specifications of future TOEFL iBT test tasks and task types and may affect the validity arguments that are made for the TOEFL iBT test.

Acknowledgments

We thank all of the individuals who contributed texts to this project. We also thank the transcribers who helped make the TMLE corpus a reality.

References

Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, *4*(1), 71 – 83. https://doi.org/10.3366/E1749503209000227

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255 – 278. https://doi.org/10.1016/j.jml.2012.11.001

Bartoń, K. (2019). *MuMIn: Multi-Model Inference* (Version 1.43.6) [Computer software]. https://cran.r-project.org/web/packages/MuMIn/index.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1 – 48. https://doi.org/10.18637/jss.v067.i01

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins. https://doi.org/10.1075/scl.23

Biber, D., & Conrad, S. (2014). *Variation in English*: Multi-dimensional studies. Routledge. https://doi.org/10.4324/9781315840888

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E., & Urzua, A. (2004). *Representing language use in the university: Analysis of the TOEFL 2000 spoken and written academic language corpus* (TOEFL Monograph Series No. TOEFL-MS-25). ETS.

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.

Brooks, L., & Swain, M. (2014). Contextualizing performances: Comparing performances during TOEFL iBT and real-life academic speaking activities. *Language Assessment Quarterly*, *11*(4), 353 – 373. https://doi.org/10.1080/15434303.2014.947532

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Brunfaut, T., & Révész, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, *49*(1), 141 – 168. https://doi.org/10.1002/tesq.168

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1 – 47. https://doi.org/10.1093/applin/1.1.1

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, *42*(3), 475–493. https://doi.org/10.1002/j.1545-7249.2008.tb00142.x

Crossley, S. A., Kyle, K., Davenport, J., & McNamara, D. S. (2016). *Automatic assessment of constructed response data in a chemistry tutor*. Paper presented at the 9th International Educational Data Mining Society Conference, Raleigh, NC.

R Development Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.r-project.org/

Explosion AI. (2018). *SpaCy language models*. https://spacy.io/models/en#en_core_web_sm

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Heck, R. H., & Thomas, S. L. (2020). *An introduction to multilevel modeling techniques: MLM and SEM approaches* (4th ed.). Routledge. https://doi.org/10.4324/9780429060274

Hox, J. J. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge. https://doi.org/10.4324/9781315650982

Hymes, D. (1974). Ways of speaking. In R. Bauman & J. Sherzer, *Explorations in the ethnography of speaking*, 433–451. Cambridge University Press. https://doi.org/10.1017/CBO9780511611810.029

Jacoby, J. (2014). The disruptive potential of the massive open online course: A literature review. *Journal of Open, Flexible, and Distance Learning*, *18*(1), 73–85.

Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel* (Research Branch Report 8-75). University of Central Florida, Institute for Simulation and Training. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. Georgia State University. http://scholarworks.gsu.edu/alesl_diss/35/

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *Modern Language Journal*, *102*(2), 333–349. https://doi.org/10.1111/modl.12468

LaFlair, G. T., & Staples, S. (2017). Using corpus linguistics to examine the extrapolation inference: A case study of a high-stakes speaking assessment. *Language Testing*, *34*(4), 451–475. https://doi.org/10.1177/0265532217713951

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Emmeans: Estimated marginal means, aka least-squares means* (Version 1.4.7) [R package]. https://github.com/rvlenth/emmeans

Loewen, S., & Gonulal, T. (2015). Exploratory factor analysis and principal components analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 182–212). Routledge. https://doi.org/10.4324/9781315870908-9

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, *115*(3), 1–47.

Naidu, S. (2013). Transforming MOOCs and MOORFAPs into MOOLOs. *Distance Education*, *34*(3), 253–255. https://doi.org/10.1080/01587919.2013.842524

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research: Effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Revelle, W. (2016). *An introduction to psychometric theory with applications in R: An overview*. Personality Project. http://personality-project.org/r/book/

Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (Part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-Level Student Papers (MiCUSP). *Corpora*, *6*(2), 159–177. https://doi.org/10.3366/cor.2011.0011

Sawaki, Y. (2013). Factor analysis. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–5). Blackwell. https://doi.org/10.1002/9781405198431.wbeal0407

Sheehan, K. M., Kostin, I., & Futagi, Y. (2008). *When do standard approaches for measuring vocabulary difficulty, syntactic complexity and referential cohesion yield biased estimates of text difficulty?* Paper presented at the 30th Annual Conference of the Cognitive Science Society, Washington, DC.

Sheehan, K. M., Kostin, I., Futagi, Y., & Flor, M. (2010). *Generating automated text complexity classifications that are aligned with targeted text complexity standards* (Research Report No. RR-10-28). ETS. https://doi.org/10.1002/j.2333-8504.2010.tb02235.x

Staples, S., LaFlair, G. T., & Egbert, J. (2017). Comparing language use in oral proficiency interviews to target domains: Conversational, academic, and professional discourse. *Modern Language Journal*, *101*(1), 194–213. https://doi.org/10.1111/modl.12385

Yan, X., & Staples, S. (2020). Fitting MD analysis in an argument-based validity framework for writing assessment: Explanation and generalization inferences for the ECPE. *Language Testing*, *37*(2), 189–214. https://doi.org/10.1177/0265532219876226

Suggested citation:

Kyle, K., Choe, A. T., Eguchi, M., LaFlair, G., & Ziegler, N. (2021). *A comparison of spoken and written language use in traditional and technology-mediated learning environments* (TOEFL Research Report No. RR-94). ETS. https://doi.org/10.1002/ets2.12329

Action Editor: John Norris

Reviewers: This report was reviewed by the Research Subcommittee of the TOEFL Committee of Examiners.

Find other ETS-published reports by searching the ETS ReSEARCHER database at http://search.ets.org/researcher/