



## Comparing writing performance in TOEFL-iBT and academic assignments: An exploration of textual features



A. Mehdi Riazi\*

Department of Linguistics, Macquarie University, Balaclava Road, North Ryde 2109, NSW, Australia

### ARTICLE INFO

#### Article history:

Received 9 November 2015

Received in revised form 25 January 2016

Accepted 3 February 2016

Available online 21 February 2016

#### Keywords:

Writing

Academic writing

Writing assessment

Validity of writing tests

Text analysis

TOEFL-iBT-Coh-Metrix

### ABSTRACT

This paper reports an exploratory study in which the written texts produced by postgraduate students in test and real-life academic situation are compared in terms of the linguistic and discursal features. Data were collected from 20 international English as a second language (ESL) postgraduate students from different first language backgrounds and three general disciplines of science and engineering, arts and humanities, and business and economics. The participants were studying in postgraduate programs in five universities in New South Wales, Australia. These participants completed two writing test tasks of the TOEFL-iBT (integrated and independent tasks) and an academic assignment for one of the university courses they enrolled in. Textual features of the test and academic assignment texts were compared on 20 linguistic and discursal features. These textual features are related to syntactic complexity (five variables), lexical sophistication (nine variables) and cohesion (six variables). Results of a series of repeated measures Analysis of Covariance (ANCOVA) indicated similarities and differences in the linguistic and discursal features of the three writing task texts. Findings are reported and discussed and implications are made for the extrapolation inference claim in the validity argument of the Writing section of the TOEFL-iBT.

© 2016 Elsevier Inc. All rights reserved.

### 1. Introduction

A perplexing question with regard to large-scale and high-stakes language proficiency tests like TOEFL is the extent to which test takers' performance on timed writing tasks in test situation can be used as an index to predict test takers' writing performance in university. The extrapolation inference in the test of English as a foreign language (TOEFL) validity argument framework (Chapelle, Enright, & Jamieson, 2008; Enright & Tyson, 2008) includes the claim that examinee performance on test tasks could be generalized to their performance on academic tasks they encounter in university settings. The extrapolation inference in the TOEFL argument framework is based on the warrant that "the construct of academic language proficiency as assessed by the TOEFL accounts for the quality of linguistic performance in English-medium institutions of higher education" (Chapelle et al., 2008, p. 21). Accordingly, based on both theoretical conceptualisations (e.g., Cumming, Kantor, Powers, Santos, & Taylor, 2000; Hamp-Lyons & Kroll, 1996) and empirical studies (e.g., Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Hale et al., 1996) a set of writing tasks were designed for the new TOEFL test so that they could allow test takers to demonstrate relevant writing performance in a manner that would provide evidence of their writing ability as

\* Corresponding author. Tel.: +61 406682439/+61 2 98507951; fax: +61 298509199.  
E-mail address: Mehdi.riazi@mq.edu.au

required in institutes of higher education. Subsequently, the new internet-based TOEFL (TOEFL-iBT) was launched in 2005 with two writing test tasks, namely, an integrated task (reading, listening, writing) and an independent task (responding to a writing prompt). [Cumming et al. \(2006\)](#) provided some evidence of discourse features and verification of scoring levels for the authenticity of the independent and integrated prototype writing tasks for the new TOEFL.

Notwithstanding the above evidential support and despite the major differences between test and real-life academic writing contexts, such as the genre of writing, time allocation, text length, resource accessibility, audience, and so on, it is useful to investigate how texts produced in the two situations might be compared in terms of text quality. The aim of this study was thus to explore how the linguistic and discursive features of the texts produced in TOEFL-iBT writing test tasks compared to those of academic writing assignments. The study was based on the assumption that, despite the differences between the two situations, participants' core language and writing proficiency can be reflected in their texts and compared across the two situations (test vs. real-life academic). This would especially be revealing when a repeated measures design is used in which the texts of the same participants in the two situations are compared. The outcomes of the study provides preliminary evidence or, otherwise, rebuttal for the claims made in the extrapolation inference in TOEFL validity argument in terms of the quality of the texts produced in the test situation. For example, the similarities in text quality across the two situations (test vs. academic assignment), can be used to argue that the test tasks are capable of eliciting the same linguistic and discursive features from the test takers, as do academic writing assignments. On the other hand, the differences in the linguistic features of the two test texts and those produced in academia would warrant further research in this area.

This study used a repeated measures analysis of covariance (RMANCOVA) to compare a group of postgraduate coursework students' writings in test and academic course situations. According to [Larson-Hall \(2010, p. 323\)](#) "research designs which incorporate repeated measures are quite desirable, as they increase the statistical power of a test." Larson-Hall goes on to say that "This is an important factor to consider in our field, where sample sizes are generally small" as is the case in the current study.

## 2. Theoretical background

The conceptualization of writing in TOEFL-iBT is based on the *TOEFL 2000* ([Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000](#)) and [Cumming et al. \(2000\)](#) frameworks, which focus on testing writing in academic settings. As conceptualized by [Cumming et al. \(2000\)](#), writing in academic settings typically involves the production of written text forms to meet certain expectations to complete course-related assignments and to display the knowledge they are acquiring. Based on this conceptualisation, the assumption then is that, if test takers' were able to perform well on TOEFL-iBT writing test tasks, they would be able to use the core writing abilities to perform discipline-specific assignments.

Considering "language use", [Cumming et al. \(2000\)](#) reviewed published empirical studies that have addressed matters of language use in adult ESL students' written texts and highlighted the following as potential evaluative criteria for assessing written texts.

- Appropriateness and range of vocabulary and idiom used
- Appropriateness and effectiveness of language used to manage discourse connections
- Accuracy, appropriacy, and range of phrase-level syntax and morphological markers; and
- Spelling, punctuation, and other orthographic or typographic conventions, and for tasks directly dependent on content from reading or listening passages, a measure of ability to use specific language elements in paraphrase, reported speech, and summation. (p. 19)

[Cumming et al. \(2006\)](#) examined if and how the text and discourse features of prototype integrated tasks developed for the new TOEFL test differed from the text and discourse features of the independent TOEFL essays. [Cumming et al.'s \(2006\)](#) study followed the premises of the text characteristics model as outlined in [Cumming et al. \(2000\)](#). Having reviewed quite a large number of publications on text characteristics, [Cumming et al. \(2006\)](#) used the following indicators in their analysis of the texts:

- Text length, operationalized as the total number of words written in a composition.
- Lexical sophistication, which was analyzed in two ways: (a) average word length and (b) type/token ratio of the number of different lexical items over the total number of words per composition.
- Syntactic complexity, which was analyzed in two ways: (a) number of clauses per T-unit and (b) words per T-unit.
- Holistic rating of grammatical accuracy as either 1 (many severe errors, often affecting comprehensibility), 2 (some errors but comprehensible to a reader), or 3 (few errors, and comprehensibility seldom obscured for a reader).
- Quality of argument structures, evaluating the claims, data, warrants, proposition, oppositions, and responses to oppositions as six elements, each rated as either 0, 1, 2, or 3.
- Orientations to source evidence expressed in each T-unit, coding for presentation of voice (as references that are either unspecified or are specified with respect to evidence from self or from other(s), with a person either identified or not, or

Download English Version:

<https://daneshyari.com/en/article/344212>

Download Persian Version:

<https://daneshyari.com/article/344212>

[Daneshyari.com](https://daneshyari.com)