Assessing Writing 25 (2015) 22–37

Contents lists available at ScienceDirect

Assessing Writing

"I must impress the raters!" An investigation of Chinese test-takers' strategies to manage rater impressions

Qin Xie *

Department of Linguistics and Modern Languages, Hong Kong Institute of Education, B4-2/F-12, NO. 10 LO Ping Road, Tai Po Market, New Territories, Hong Kong

ARTICLE INFO

Article history:
Received 17 May 2014
Received in revised form 9 May 2015
Accepted 11 May 2015

Keywords:
Test-taker perception
Test-wiseness strategy
Holistic scoring
Essay tests
Test validity
Washback

ABSTRACT

Most studies on holistic scoring procedures adopt a rater perspective, focusing on raters and textual features; few studies adopt a test-taker perspective. This study investigated test-taker perceptions of impression management strategies for holistically scored essay tests and estimated the extent to which such perceptions can predict essay scores. A total of 886 Chinese test-takers took two essay tests and completed a perception questionnaire, where they rated the importance of impression management strategies versus the target language and writing skills for helping them achieve good essay scores. Four raters marked the essay papers holistically; each essay paper received two independent ratings. Questionnaire items were analysed to verify scale reliability and to explore the underlying structure of each perception scale. The analysis identified two opposing factors underlying test-takers' perceptions of the impression management strategies, which represented a risk-taking approach and a defensive approach to managing rater impressions. Regression analyses found the defensive approach consistently predicted essay scores, i.e. test-takers assigning higher values to the defensive approach achieved significantly higher scores than test-takers who assigned it lower values. Corpus-based textual analyses found defensive writers writing longer essays and committing fewer linguistic errors, while risk-takers used slightly more sophisticated words and sentences but also committed more errors. The findings were discussed in terms of their implications for validity and washback.

© 2015 Elsevier Inc. All rights reserved.

* Tel.: +852 2948 8368.
E-mail address: qxie@ied.edu.hk

http://dx.doi.org/10.1016/j.asw.2015.05.001
1075-2935/© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Holistic scoring procedures (Williamson, 1993) have been widely used both in L1 (White, 1995) and L2 writing assessment (Weigle, 2002) in close association with timed impromptu essay tests. Existing research has established the reliability of holistic scoring, yet its impressionistic nature has long been a source of concern for assessment researchers (Barkaoui, 2011; Charney, 1984; Cumming, 2009; Hamp-Lyons, 1995). Studies on both L1 and L2 writing assessment observed that raters operating on holistic scoring scales tend to be influenced by surface textual features, such as essay length, spelling errors and handwriting. Most of these studies, however, rely on the rater perspective, focusing on reliability issues associated with raters and their rating processes. Few studies have taken the test-taker perspective and considered possible effects of test-taker perceptions on their writing process, text features and essay scores. Adopting a test-taker perspective may provide a new angle from which to examine holistic essay scores and scoring and to address important questions about score validity and washback on teaching and learning. This study investigated the links between test-taker perceptions and their holistic essay scores, and explored the textual features underlying these relationships. The study is situated within the education context of China.

Given the gigantic scale of China's test-taker population (Cheng & Curtis, 2010), it is not surprising that holistic scoring dominates the marking sites of important, large-scale English proficiency or achievement tests. According to Qi (2007), analytical scoring was implemented only for a short period during the earlier years of the National Matriculation English Test (NMET), and was later replaced by the holistic scoring method in 1992. Hence, test-takers from Mainland China have very limited exposure to alternative methods of essay scoring. This study used a questionnaire to collect test-taker perceptions regarding a specific essay test within a large-scale English language proficiency test in China (i.e. College English Test Band 4, CET4). Participants' essay papers and their official scores on CET4 were also collected. The relations between test-taker perceptions and their holistic essay scores were estimated statistically. Afterwards, essay papers were sampled and analysed to identify textual features that explained the relations between test-taker perceptions and their essay scores.

2. Literature review

2.1. Research on holistic scoring and essay tests

Despite the advent of portfolio assessment, timed impromptu essay tests are still widely used, especially in large-scale second language proficiency tests, to assess test-takers' ability to write in the target language. Essay papers are still scored primarily by human raters adopting holistic scoring procedures, although recent developments in computer-assisted automated essay scoring (Lee, Gentile, & Kantor, 2010; Ramineni & Williamson, 2013) may change the status quo in the near future. Holistic scoring is described as a set of quick impressionistic procedures for assigning scores to a written work according to pre-established criteria (White, 1995; Charney, 1984). On the basis of their overall impression of a paper, essay raters assign a grade to an essay by mapping it against an ordered scale of descriptors (Charney, 1984). Because judgement is usually made in a matter of seconds or minutes, scores assigned are largely based on first impressions which may convey more about the appearance of an essay than its substance. Thus, the validity of the essay scores obtained holistically has been a source of concern for assessment researchers for decades (Barkaoui, 2011; Caudery, 1990; Charney, 1984; Hamp-Lyons, 1995; Huot, 1990).

Existing studies (Bacha, 2001; Gardiner, 2004; Lim, 2011; Weigle, 2002; Williamson, 1993) found that raters utilizing holistic scoring procedures can achieve a high level of reliability through training and marking experience. Studies focusing on validity (e.g. Barkaoui, 2010; Gamaroff, 2000; Lumley, 2002), however, found that holistic scoring procedures allow raters to deviate from the intended assessment criteria to their own implicit evaluative criteria. They may prioritize different rating criteria and assign different weights to different components of the criteria. Thus, similar scores may not suggest similar judgement, while different scores may not indicate different judgement.

Similarly, the studies on the textual features of essays (written either in L1 or L2) found that surface textual features, such as essay length, handwriting, spelling errors, word choice and range, and

Download English Version:

https://daneshyari.com/en/article/344207

Download Persian Version:

https://daneshyari.com/article/344207

Daneshyari.com