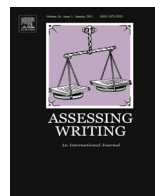




ELSEVIER

Contents lists available at [ScienceDirect](#)

Assessing Writing



Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks



Atta Gebril^{a,*}, Lia Plakans^b

^a Department of Applied Linguistics, The American University in Cairo, P.O. Box 71, 11835 New Cairo, Egypt

^b Lia Plakans, Department of Teaching and Learning, N259 Lindquist Hall, Iowa City, IA 52242, USA

ARTICLE INFO

Article history:

Received 12 November 2013

Received in revised form 26 February 2014

Accepted 3 March 2014

Available online 22 March 2014

Keywords:

Writing assessment

L2 writing

Integrated tasks

Reading-to-write

Validity

Scoring

ABSTRACT

Integrated writing tasks that depend on input from other language abilities are gaining ground in teaching and assessment of L2 writing. Understanding how raters assign scores to integrated tasks is a necessary step for interpreting performance from this assessment method. The current study investigates how raters approach reading-to-write tasks, how they react to source use, the challenges they face, and the features influencing their scoring decisions. To address these issues, the study employed an inductive analysis of interviews and think-aloud data obtained from two raters. The results of the study showed raters attending to judgment strategies more than interpretation behaviors. In addition, the results found raters attending to a number of issues specifically related to source use: (a) locating source information, (b) citation mechanics, and (c) quality of source use. Furthermore, the analysis revealed a number of challenges faced by raters when working on integrated tasks. While raters focused on surface source use features at lower levels, they shifted their attention to more sophisticated issues at advanced levels. These results demonstrate the complex nature of integrated tasks and stress the need for writing professionals to consider the scoring and rating of these tasks carefully.

© 2014 Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +20 226151919; fax: +20 227957565.

E-mail addresses: agebril@aucegypt.edu (A. Gebril), Lia-plakans@uiowa.edu (L. Plakans).

1. Introduction

Integrated writing tasks, in which writers are required to synthesize information from external sources in their written product, are increasingly used in the context of L2 writing assessment (Gebril & Plakans, 2009, 2013). Researchers argue that such tasks ingeniously replicate the actual practices in academic contexts where discourse synthesis is a common exercise in university writing (Gebril, 2009; Hale et al., 1996; Horowitz, 1986; Moore & Morton, 1999; Plakans, 2008, 2009). According to the results of these studies, most university professors require students to write in a variety of genres that need a considerable amount of source-based writing, such as research papers, reports, reaction papers, and case studies. Hence, integrated assessment methods are expected to augment authenticity and better elicit the academic writing construct.

Integrated tasks are “more complex and more demanding than traditional stand-alone or *independent* tasks, in which test-takers draw on their own knowledge or ideas to respond to a question or prompt” (Brown, Iwashita, & McNamara, 2005, p. 1). This complexity is not limited to task performance, but also to scores on integrated tasks. Generally, writing scores have relatively low reliability given the different construct-irrelevant variables confounding writing scores, and integrated writing tasks are no exception (Gebril, 2009, 2010). Consequently, teachers and language testers have constantly been concerned about the use of integrated tasks in their writing exams. Discussions usually arise among writing practitioners about rating-related issues, such as assessment criteria, discourse synthesis, and ownership. Critical in these debates are questions pertaining to how raters or instructors¹ go about assigning scores to integrated tasks, what criteria they should use, and what constitutes appropriate textual borrowing practices. Given the centrality of these issues, development of valid and reliable assessment criteria for scoring integrated writing tasks is urgently needed to ensure their effective implementation in writing courses and programs.

Documenting the different processes raters use when scoring written responses is of strategic importance to building sound validity evidence for writing tests. According to current thinking in the assessment field, scoring validity must be thoroughly investigated as part of test validation (Kane, 2006; Shaw and Weir, 2007). Issues of interest in scoring validity investigation include assessment criteria used by raters, rater agreement, score reliability, and scale accuracy. Although there is growing interest in rater cognition in writing assessment research (e.g. Barkaoui, 2011; Cumming, Kantor, & Powers, 2001, 2002; Lumley, 2002; Milanovic, Saville and Shuhong, 1996; Weigle, 1994), little research has investigated what raters attend to when scoring second language writing and whether resulting scores reflect actual writing proficiency (Cumming et al., 2001). Douglas and Selinker, 1992 argue that raters may assign similar ratings for totally different reasons in spite of the fact that they are using the same scoring rubric. In another study, Douglas, 1994 concluded that scoring discrepancy could be due to discourse features that were not included in the scoring rubric, and recommended using think-aloud studies of rater cognition for better understanding of how raters arrive at a certain score, which is a methodology used in the current study.

2. Literature review

One strand of writing assessment research has addressed the criteria that raters use when scoring L2 writing tasks (e.g. Cumming, 1990; Cumming et al., 2001; DeRemer, 1998; Lumley, 2002). These studies depended mainly on the think-aloud methodology, with raters verbalizing the processes they use when scoring a written text. For example, Lumley, 2002 collected think-aloud data from four raters while scoring writing samples as part of an Australian proficiency test. The results of his study suggested that the raters used similar strategies during the scoring process: reading the essay, assigning a score, and then reconsidering this assigned score. However, the data also revealed that the relationship between the writing sample and the scoring rubric remains obscure. Consequently, the raters in the study came up with unique strategies to accommodate this problem. Lumley argued that the

¹ In L2 writing settings, instructors often work as raters. Throughout this article, we will use the term raters as inclusive of teachers who are rating students' work as well as non-instructor raters.

Download English Version:

<https://daneshyari.com/en/article/344229>

Download Persian Version:

<https://daneshyari.com/article/344229>

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