



Investigating how vocabulary is assessed in a narrative task through raters' verbal protocols



Hui Li ^{b, *}, Nuria Lorenzo-Dus ^a

^a College of Arts and Humanities, Swansea University, Singleton Park, Swansea SA2 8PP, Wales, UK

^b Department of English, Culture and Communication, Xi'an Jiaotong-Liverpool University, No.111 Ren'ai Road, Dushu Lake Higher Education Town, Suzhou Industrial Park, Jiangsu Province 215123, PR China

ARTICLE INFO

Article history:

Received 9 July 2012

Received in revised form 21 May 2014

Accepted 30 June 2014

Available online

Keywords:

Rater protocols

Vocabulary assessment

Analytic rating scales

Oral proficiency

ABSTRACT

The study examines the process of assessing vocabulary in oral proficiency examinations. Vocabulary is increasingly adopted as an effective indicator of candidates' oral proficiency in large-scale tests, but there is limited empirical evidence so far regarding how raters assess it. In this experiment, 25 participants rated one English oral text produced by a candidate with Chinese as a first language. Raters' verbal protocols were transcribed and coded to identify what raters attended to in assessing vocabulary. The candidate's use of 'advanced' words was found to have a direct impact on vocabulary scores. Also, both vocabulary and non-vocabulary features emerged in the raters' protocols. The findings question the possibility of assessing vocabulary as a discrete construct.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Research on oral proficiency in a second language shows a positive correlation between learners' vocabulary and their reading, writing, listening and speaking skills (e.g. Brown, Iwashita, & McNamara, 2005; Luoma, 2004). In both replacement and high-stakes tests, for instance, vocabulary is often assessed as an independent construct, i.e., vocabulary scores are treated either as a discrete aspect of candidates' language profile or as a distinctive contributor to a composite final score. The 2001 revised version of IELTS speaking test, for instance, included lexical resource (vocabulary) as one of four analytical subscales to be assessed: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation (Brown & Taylor, 2006). The assumption is that vocabulary could be an indicator of candidates' oral proficiency and a construct that can be assessed discretely.

However, little is known about *how* vocabulary may be assessed as an independent construct in oral proficiency examinations. Therefore, the main aim of this study is to investigate the process of assessing vocabulary in oral proficiency contexts. In order to do so, we examined how 25 individuals rated the same oral text – a narrative – produced by one candidate, an intermediate-level learner of English as a foreign language. This study sought to identify (i) what vocabulary features are most/least frequently attended to by raters and (ii) what impact these features have on raters' scores.

2. Assessing spoken vocabulary in a foreign language

The importance of vocabulary knowledge is well-documented in the literature. Vocabulary size, for instance, is closely related to reading comprehension (Nation, 2001; Schmitt & McCarthy, 1997), writing ability (Baba, 2009; Engber,

* Corresponding author. Tel.: +86 (0)512 8816 1307.

E-mail addresses: Hui.Li@xjtlu.edu.cn (H. Li), n.lorenzo-dus@swansea.ac.uk (N. Lorenzo-Dus).

1995), listening comprehension (Bonk, 2000; van Zeeland & Schmitt, 2013) and, crucially for the present study, oral proficiency (Iwashita, Brown, McNamara, & O'Hagan, 2008; Read, 2000, 2005). Spoken vocabulary, in turn, is regarded by many as an effective construct in foreign language proficiency placement tests (Adams, 1980; Laufer & Nation, 1999; Lorenzo-Dus & Meara, 2005). Indeed, for Iwashita (2010) and Lu (2005), vocabulary is one of the best indicators of oral proficiency level.

Furthermore, vocabulary has been identified as a salient feature attended to by raters in speaking tests (Zhang & Elder, 2011), but conflicting views on the judgement of vocabulary also surfaced. For example, when examining raters' performance in TOEFL oral examinations, Brown et al. (2005) found that vocabulary was frequently mentioned by raters when explaining how they arrived at their scores. Meanwhile, the authors also found that their raters displayed conflicting views regarding which aspects a given candidate's lexical performance was considered 'most important': some raters favoured the use of academic vocabulary; whereas others regarded colloquial vocabulary as a key indicator of oral proficiency. In a study by Lorenzo-Dus and Meara (2005) on the relationship between rater accommodation strategies and test-taker vocabulary in oral proficiency examinations, 'rarity of vocabulary' emerged as a particularly salient factor in determining scores in the raters' protocols, and statistically significant correlations were found between lexical diversity (VocD, the number of different words used) and oral scores. Likewise, in Malvern and Richards' (2002) analysis of oral proficiency interviews, candidates' use of 'less common words' was suggested as an especially noteworthy feature for raters, but unlike Lorenzo-Dus and Meara (2005), no statistically significant correlation was found between lexical diversity (VocD) and vocabulary scores. Past studies have also suggested that, in assessing spoken vocabulary, raters typically make more negative than positive comments (Brown, 2006). Therefore, how vocabulary is judged by raters merits attention and further empirical studies.

While useful in identifying the roles that different vocabulary features may play in the assessment of spoken vocabulary, analyses that correlate lexical features and candidates' scores – such as those reported above – do not focus on the decision making processes that lead to the award of specific vocabulary scores in oral proficiency interviews. Analytic methods devised to tease out these processes – i.e., how raters assess spoken vocabulary, are thus also needed. Verbal protocol analysis (VPA) constitutes one such data-gathering method and is the one used in our study.

In VPA, individuals are given a task (e.g. solving a problem) and are asked to 'vocalize what is going through their minds as they are solving a problem or performing a task' (Gass & Mackey, 2000, p. 13). VPA is based on the assumption that an individual's verbalisation may be perceived as an account of information that is attended to in the process of task performance (Falvey & Shaw, 2006). It is most commonly used to investigate the cognitive process of task-solving, and it can be collected either concurrently (during an event) or retrospectively (after an event). Unlike some observational methods that focus on linguistic components or structures, such as discourse and conversation analytic techniques, VPA examines the information therein to make inferences about the cognitive process of task performance. It is seen as a quantitative approach to analysing qualitative data. It helps to count and identify the relationship between contents of utterances in order to mitigate the subjectiveness of interpretations of qualitative data (Chi, 1997).

While it is acknowledged that the verbalisation of cognitive processes through verbal protocols may not accurately and/or fully capture decision-making processes, their validity and reliability have been thoroughly tested (Cohen, 1987; Fulcher, 2003; Gass & Mackey, 2000). By and large, VPA is accepted as a viable data analysis method (see Green, 1998 for a full account of the methodology; Swain, 2006) to investigate raters' decision-making processes when assessing writing (Cumming, Kantor, & Powers, 2002) and speaking (Ducasse & Brown, 2009; Nakatani, 2010; Yates, Zielinski, & Pryor, 2011). A potential weakness of using verbal-report methods to investigate speaking rather than writing is that 'the real-time nature of the assessment ... limits what can be inferred about the process of rating, as opposed to the performance features to which raters attend (Brown et al., 2005, p. 7)'. Despite this, VPA is a popular method in tracing raters' decision-making processes in oral contexts.

In the field of language testing, verbal protocols have been used to analyse both raters' and test-takers' performances. Regarding the former, VPA has been adopted to trace decision-making processes in assessing compositions (Barkaoui, 2011; Cumming et al., 2002; Milanovic, Saville, & Shen, 1996); interpretation of rating scales (Falvey & Shaw, 2006); and how training and background influence raters' behaviour (Hamp-Lyons, 1990; Weigle, 1994, 1999). The results of the few available studies of oral proficiency assessment using VPA coincide in concluding that examiners apply rating scales to candidates inconsistently and attend to some facets of candidates' performance that are not stated in the scales (Ang-Aw & Goh, 2011; May, 2006; Orr, 2002). In Brown et al. (2005), raters' perceptions of candidates' oral proficiency were found to be principally driven by language content, 'traditional linguistic resources' (i.e., grammar and vocabulary) and production features (i.e., fluency and pronunciation). Within vocabulary, their VPA revealed that raters often focussed on accuracy, lexical sophistication, range (lexical diversity), use of collocations, and appropriateness of specific lexical choices. Their study, however, did not report convincing statistical evidence regarding which vocabulary features were most salient; no detailed analysis of rater variation in assessing any of these features was provided either.

In light of the above, our study examines the process of assessing spoken vocabulary by addressing two research questions:

Research question 1: what features do raters claim to attend to when assessing spoken vocabulary?

Research question 2: what impact, if any, do the specific vocabulary features attended to by raters have on their scores?

Download English Version:

<https://daneshyari.com/en/article/6849553>

Download Persian Version:

<https://daneshyari.com/article/6849553>

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