



# “I can see that”: Developing shared rubric category interpretations through score negotiation



Jonathan Trace<sup>a</sup>, Valerie Meier<sup>b,\*</sup>, Gerriet Janssen<sup>c</sup>

<sup>a</sup> Keio University, Faculty of Environment and Information Studies, 5322 Endo, Fujisawa City, Kanagawa, Japan

<sup>b</sup> University of California, Santa Barbara, Gevirtz Graduate School of Education, Santa Barbara, CA 93106-9490, USA

<sup>c</sup> Universidad de los Andes, Depto de Lenguas y Cultura, Carrera 1 No. 18A-12 Edificio Franco, 5° piso, Bogotá, Colombia

## ARTICLE INFO

### Article history:

Received 1 November 2015

Received in revised form 24 June 2016

Accepted 5 August 2016

Available online 25 August 2016

### Keywords:

Academic writing

Negotiation

Performance assessment

Rubrics

Sociocultural learning

Raters

## ABSTRACT

Performance assessments using raters will always contain some subjectivity, and disagreement among raters necessitates reliable methods for resolving scores. Negotiation is one effective method to guide scoring decisions and reduce raters' tendencies to be unexpectedly severe or lenient when scoring specific rubric categories or examinees. Beyond its utility for scoring, however, negotiation is also a resource for raters to co-construct interpretations about the language constructs being measured.

This study uses quantitative and qualitative methods to trace how negotiation impacts raters' scoring decisions and examine in detail how raters develop joint interpretations of rubric category criteria. Scores from the writing section of a high stakes English language placement exam ( $n = 60$ ) were analyzed using ANOVA and many-faceted Rasch measurement to determine which categories were frequently assigned discrepant scores and to estimate rater severity. Discourse analysis of six audiotaped negotiation sessions was then used to examine how raters' understanding of rubric criteria converged over time. Our results indicate that through negotiation, raters used shared terminology and justifications to clarify ambiguous constructs and work to establish shared values. The results suggest that score negotiation influences scoring inferences and also creates affordances for raters to ground those inferences in shared constructions of meaning.

© 2016 Elsevier Inc. All rights reserved.

## 1. Introduction

Research on raters in writing performance assessment has generally focused on rater reliability and rubric function (i.e., rubric category validity). Studies have looked at raters in terms of score reliability and validity (Bachman & Lynch, 1995; Kondo-Brown, 2002; Lim, 2011; Lumley & McNamara, 1995; Weigle, 1999), while other researchers have considered the reliability and validity of rating scales themselves (East, 2009; Upshur & Turner, 1995). Little has been written, however, about the process leading up to the final scores, particularly the use of negotiation to resolve discrepant scores.

In writing performance assessments, it is common to use multiple raters to score a single test taker's performance. However, research has shown that even trained raters exhibit significant differences in their overall level of severity and (Bonk & Oakey, 2003; Eckes, 2005; Lumley & McNamara, 1995; Weigle, 1998; Wigglesworth, 1993). Consequently, programs using performance assessments must account for the fact that raters are likely to assign discrepant scores and have in

\* Corresponding author.

E-mail addresses: [tracej@sfc.keio.ac.jp](mailto:tracej@sfc.keio.ac.jp) (J. Trace), [vmeier@education.ucsb.edu](mailto:vmeier@education.ucsb.edu) (V. Meier), [gjanssen@uniandes.edu.co](mailto:gjanssen@uniandes.edu.co) (G. Janssen).

place a principled approach to address this situation. Negotiation is one strategy that has been used to resolve scoring discrepancies and minimize the influence of construct-irrelevant score variability (e.g., Johnson, Penny, & Gordon, 2000; Johnson et al., 2001; Johnson, Penny, Gordon, Shumate, & Fisher, 2005). Although the effects of negotiation have been studied from psychometric perspectives (e.g., Trace, Janssen, & Meier, 2015), the interaction process itself has been somewhat neglected. While Trace, Janssen, and Meier reported that raters, through the discussion of discrepant scores, began to have a more sophisticated understanding of rubric constructs, the study did not describe the characteristics of these interactions. Therefore, the present study focuses on exploring in more detail the interactions that take place during score negotiation.

Sociocultural and ecological perspective on learning and interaction provide one productive perspective from which to explore these interactions. While this approach has been discussed in the assessment literature, primarily in the areas of mediated or dynamic assessment (e.g., Anton, 2009; Lantolf & Poehner, 2010; Poehner, 2007; Poehner & Lantolf, 2005), it also lends itself to research questions concerning performance assessment rating, as it focuses on how interaction is influenced and shaped by the physical and social environment. Consequently, in this study we adopt this framework in order to describe the ways in which raters engaged in score negotiation developed a more refined understanding of the language constructs being measured and established shared language and values.

## 2. Literature review

### 2.1. Performance assessment rating

According to current models of argument-based validity (e.g., Kane, 2012, 2013), score assignment is part of the *scoring inference*. In this step of the validation argument, test developers must gather evidence to show that “the scoring procedures are appropriate, are applied as intended, and are free of overt bias” (Kane, 2013; p. 25). In performance assessment contexts, this means demonstrating that the raters’ behavior evidences these three aspects of scoring. Kane (2006) describes *appropriateness* as the degree to which expert raters can determine which scoring categories best reflect the construct assessed. To determine if these scores are *applied* as intended, experts need to review rater training procedures, scoring guides, and scoring procedures. Lastly, *overt bias* can be checked through the use of empirical methods such as statistical analysis (Kane, 2006; p. 34). Rater bias is a major concern in the validation argument, as human raters do not mechanically apply rating scale criteria; instead, they bring to the scoring task their own subjectivities and interpretations of rubric categories and language constructs. Thus, it is not surprising that multiple raters will not assign identical scores to the same test record; previous research has thoroughly documented that even trained raters may exhibit significant differences in severity on different analytic rubric categories (Eckes, 2005; Lim, 2011; Lynch & McNamara, 1998; Schaefer, 2008; Wigglesworth, 1993).

When raters assign discrepant scores, several methods of score resolution are available. A number of these methods incorporate scores from a third rater in order to arbitrate scoring differences; however, including a third rater is not always feasible. Alternatively, the two original raters can negotiate their initial scores, discussing the essay, scoring rubric criteria, and benchmark essays in order to arrive at more congruent scores (Johnson et al., 2005). Like other score resolution methods, however, this process may jeopardize the scoring inference in the exam’s validation argument. This has been the focus of previous research. Johnson et al. (2005) concluded that in their context achieving a consensus score through negotiation did not improve score accuracy, or the degree to which negotiated scores were aligned with expert-based criterion scores. Trace, Janssen, and Meier (2015), however, found that negotiation almost entirely eliminated significantly biased interactions between raters and specific rubric categories and between raters and individual test takers, enhancing scoring consistency. Although the study of the conditions of negotiation and how these affect the validity of the scoring inference deserve further empirical attention, in this paper we turn to a different important potential outcome of negotiation: the creation of meaning in relation to the language constructs being measured.

### 2.2. Sociocultural theory

Underdeveloped in the conversation on score negotiation is the value that negotiation may have as a means of creating shared meaning. Sociocultural or ecological meaning-making perspectives look at language as an activity that is necessarily situated in relation to the environment. These perspectives can contribute to our understanding of negotiation, as score negotiation is also a highly contextualized process, typically involving a specific language program, assessment tools with specific intended uses or interpretations, and individual raters who bring different experiences, personalities, interpersonal relationships, and interpretations of language performance and ability.

Sociocultural theory (SCT) contends that mental activity and behavior are shaped through our relationships with the surrounding social, cultural, institutional, and historical environments (Lantolf & Thorne, 2000, 2006; Wertsch, del Río, & Alvarez, 1995). Accordingly, any activity cannot be understood apart from its context. In other words, people’s thoughts and behaviors are influenced by where they are, the others around them, the experiences that brought them to that place, the physical objects and signs that fill their environment, and the interaction between these different components. This interaction is accomplished through *mediation*, which involves the use of tools to accomplish activities. Language is one tool people use to mediate the activities of expressing opinions, desires, values, and so forth. Language must be understood in relation to the context surrounding its use; while language in isolation can be rule-bound and make sense, it has little meaning unless it is situated within a context. SCT also considers *development* and *internalization*, and these are often framed

Download English Version:

<https://daneshyari.com/en/article/4935800>

Download Persian Version:

<https://daneshyari.com/article/4935800>

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