TITLE

Improving interrater reliability in the candidate selection process

ABSTRACT

This session demonstrates a class exercise created to teach management students how to improve interrater reliability when rating candidate applications. Qualtrics was used to distribute a survey instrument designed to improve rater focus on fit with position qualifications and minimize implicit bias. Students rate multiple candidates using the instrument, and reports are delivered to show the degree evaluations of candidates differed between raters. Students learn the value of creating rigorous rubrics to improve the effectiveness of the selection process. Participants in this session will practice using the instrument and discuss applications in management and human resource management classes.

KEYWORDS

Applicant selection, interrater reliability, human resource management

INTRODUCTION

After completing a job analysis, crafting a job description, and choosing when and where to post a job, hiring committees must decide which candidates to invite for interviews. Courses that teach human resource management functions, like recruitment and selection, are likely to focus on how to minimize implicit bias and prejudice in the process that leads up to hiring a candidate. For example, textbooks and best practices encourage decision-makers to use structured behavioral interview questions that are scored using a detailed rubric (see for example, Bauer, Erdogan, Caughlin, & Truxillo, 2020). Questions that are based on the qualifications outlined in the job description can minimize the opportunity for implicit bias to make its way into scoring. The use of detailed rubrics and descriptions ensure that, no matter who is rating the candidate, raters are likely to score the candidate similarly to other raters. When raters score candidates similarly, the method is considered to have interrater reliability (consistency between raters, e.g., Viswesvaran, Ones, & Schmidt, 1996). Thus, one topic we discuss in our upper-division undergraduate management classes is how to create selection rubrics that guide raters to make consistent connections to qualifications rather than feelings. In class, we speak about interrater reliability as a correlation statistic that may indicate flaws in a method or encourage confidence in a choice. Our conference session will describe the exercise in detail, provide participants (either virtual or face-to-face) an opportunity to use the rubric, test their own rating accuracy and consistency, and allow an opportunity for discussion of application of this exercise in a variety of management or decision-making courses.

THEORETICAL FOUNDATION AND TEACHING IMPLIATIONS

A resume or vitae is often included when an applicant applies for a position. It includes qualification information, but since there is no one standard layout or content, raters may weigh aspects like grade point average, personal statements, or style very differently (Schmidt & Hunter, 1998; Thoms, McMasters, Robers, & Dombkowski, 1999). The literature in performance review ratings (Viswesvaran, Ones, & Schmidt, 1996) and structured interview ratings (Campion, Palmer, and Campion, 1997; Chapman & Zweig, 2005; Hartwell & Campion, 2016) suggest to students and managers methods to improve the predictive validity and interrater reliability of resume ratings to choose who a committee will interview (who reaches the next step in the selection process). With more structure in rating scales (e.g. detailed anchors/descriptions of what is desired) and rater accuracy training, raters can focus their attention on criteria sought regardless of the varying formats or styles candidates use to design their application documents.

In the following sections, we describe a class exercise which offers students the opportunity to practice using a structured rubric to make distinctions between job candidates. This type of active learning exercise directly applies to professional roles, thus it provides students a rigorous and relevant experience they can use when making future decisions. We created an in-class exercise to teach upper division undergraduate students about interrater reliability by having them use a structured rubric to rate cover letters and resumes. We have used this exercise in required management course entitled “Principles of Management” and have used it in both fully remote classes and face-to-face classes.

LEARNING OBJECTIVES

The learning objectives for this exercise focus on process and practice.

1. Students identify benefits and drawbacks of using applicant cover letters and resumes as methods for selection.
2. Students complete in-class training before evaluating applications and practice rating candidates using a structured rubric focused on qualifications and fit.
3. Students understand and identify implicit biases and perceptual errors that contribute to poor interrater reliability scores in the appraisal of cover letters and resumes.
4. Students reflect on methods to improve rating instruments and rater accuracy training.
5. Students suggest improvements upon a process of candidate differentiation.

EXERCISE OVERVIEW

For this exercise, students assume the role of recruiter. They are each presented with two candidates to review for an internship posting. Each candidate provides a cover letter and a resume. Recruiters use a Qualtrics survey/rubric we created to help them differentiate between candidates based on fit with position qualifications rather than feelings, biases, stereotypes, or prejudice. The first two semesters we used this exercise, we gave students the opportunity to rate two of their classmates’ cover letters and resumes for a common job posting (students also practice cover letter and resume writing in this course). In the spring 2022 semester the entire class will review two common non-peer candidates. Comparisons between semesters will be shared during the session.

Each student scores the application documents to determine which candidate to interview. Candidate evaluation reports from Qualtrics are distributed for analysis and discussion. In a full class debrief and reflection, the recruiters are shown how well their rating agreed with the other recruiters’ ratings. When there are significant differences between the scores, students are guided through discussion in how to investigate issues with the instrument/the rubric, the rating training, or rater accuracy (it becomes obvious in the reports who put in the effort to review the candidate versus who did not). Recruiters receive feedback on their performance and discuss the benefits of an evaluation survey that focuses on qualification fit. They end the experience by making recommendations regarding improvements that can improve the rubric and decision-making.

SESSION DESCRIPTION

The conference session would ideally be held within a 60-minute block to allow participants to rate candidate applicant packets. Interrater reliability scores can be computed during the session and a sample report can be displayed within the session.

* 0-10 minutes: Description of course and the exercise.
* 10-30 minutes: Participants evaluate cover letters and resumes for a job posting.
* 30-40 minutes: While the results combine, we will discuss our experiences with the exercises including the benefits and drawbacks.
* 40-50 minutes: We will show a sample report and sample reliability scores. Discussions will focus on why scores agreed or differed.
* 50-60 minutes: Discussion of how a similar design could work in other management or organizational behavior courses.

REFERENCES

Bauer, T., Erdogan, B., Caughlin, D., & Truxillo, D. (2020). *Human Resource Management: People, Data, and Analytics.* Sage.

Campion, M.A., Palmer, D.K., & Campion, J.E. (1997). A review of structure in the selection interview. *Personnel Psychology,* 50, 655-702.

Chapman, D.S. & Zweig, D.I. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *Personnel Psychology,* 58, 673-702.

Hartwell, C.J. & Campion, M.A. (22016). Getting on the same page: The effect of normative feedback interventions on structured interview ratings. *Journal of Applied Psychology,* 101, 757-778.

Schmidt, F.L. & Humter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin,* 124, 262-274.

Thoms, P., McMasters, R., Roberts, M.R., and Dombkowski, D.A. (1999). Resume characteristics as predictors of an invitation to interview. *Journal of Business and Psychology,* 13(3): 339-356.

Viswesvaran, C., Ones., D.S., & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology,* 81*,* 557-574.