Will the Number of Behavioral Anchors Increase Rater Accuracy in Performance Evaluations?

Kelsey Byerly & Mark Frame, Ph.D.
Middle Tennessee State University

Abstract

Purpose: To investigate whether defining more anchors on behaviorally anchored rating scales increases rater accuracy in performance evaluations.

Research has found that behaviorally anchored rating scales (BARS) yield more accurate ratings when compared to other scales. However, research has yet to explore the intended effects of including more or less anchors using BARS format regarding psychometric properties and overall accuracy. We propose that defining five anchors on BARS will produce more accurate results compared to only defining three anchors. In addition, we will explore whether the level of rater conscientiousness relates to overall accuracy. Participants will be recruited from Middle Tennessee State University. Participants will be assigned to evaluate the performance of University professors using either BARS with five behavioral anchors or three behavioral anchors.

Introduction

Performance evaluations assess how well an individual is performing on the job by using a rating form. One standardized method that has been found to yield accurate results is behaviorally anchored rating scales (Burnaska and Hollmann, 1974; Campbell et al., 1973). BARS consist of examples of possible behaviors that might be observed on the job. The number of anchors on BARS may increase the accuracy of the instrument, and consequently provide more effective performance feedback. In addition, conscientiousness is related to overall performance in a variety of jobs, so it is likely to influence how individuals carry out their job tasks and responsibilities such as conducting performance evaluations. Ogunfowora, Bourdage, and Lee (2010) found a negative relationship between rater conscientiousness and overall performance ratings.

Hypotheses and Research Questions

Hypothesis 1: BARS with five examples will be more accurate in scoring performance evaluations than BARS with three examples.

Hypothesis 1a: BARS with five examples will result in higher scale alphas on performance evaluation ratings than BARS with three examples.

Hypothesis 1b: BARS with five examples will result in lower absolute differences between on performance evaluation ratings and true score estimates than BARS with three examples.

Hypothesis 1c: BARS with five examples will result in raters having less over rating and under rating (as compared to true score estimates) on their performance evaluation ratings than BARS with three examples.

RQ 1: Are there differences in the scale alphas of the ratings provided based upon the degree of rater conscientiousness?

RQ 1a: Are there differences in the scale alphas among high conscientious raters on performance evaluation ratings evaluated with BARS with three examples compared to BARS with five examples?

RQ 1b: Are there differences in the scale alphas among low conscientious raters on performance evaluation ratings evaluated with BARS with three examples compared to BARS with five examples?

RQ 2: Do accuracy estimates (absolute value and over-under ratings) vary for ratings provided based upon rater conscientiousness?

RQ 2a: Do accuracy estimates (absolute value and over-under rater) vary among high conscientious raters on performance evaluations evaluated with BARS with three examples compared to BARS with five examples?

RQ 2b: Do accuracy estimates (absolute value and over-under rater) vary among low conscientious raters on performance evaluations evaluated with BARS with three examples compared to BARS with five examples?

BARS with Three Anchors

Please rate the professor's ability to use visual aids.

BARS with Five Anchors

Please rate the professor's ability to use visual aids.

Participants and Materials

Participants will be recruited from Middle Tennessee State University. Approximately 120 students will be recruited.

Method

The study will include videos of University professors lecturing about a topic for an undergraduate course. All participants will receive a short training on BARS. Participants will be randomly assigned to evaluate the performance of professors using either BARS with five behavioral anchors or three behavioral anchors. Each participant, regardless of condition, will rate the professors on performance dimensions and provide an overall rating. In addition, all participants will complete the HEXACO.

Proposed Analyses

In order to test the hypotheses, a between-subjects design will be conducted.

References

Burnaska, R. F., & Hollmann, T. D. (1974). An empirical comparison of the relative effects of rater response biases on three rating scale formats. *Journal of Applied Psychology*, *59*(3), 307–312.

Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. (1973). The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology*, *57*(1), 152 22.

Ogunfowora, B., Bourdage, J., & Lee, K. (2010). Rater personality and performance dimension weighting in making overall performance judgments. *Journal of Business and Psychology*, *25*(3), 465.