Effects of training intensity on observers' ratings of anxiety, social skills, and alcohol-specific coping skills

Abigail C. Angkaw[a], Giao Q. Tran[a,*], David A.F. Haaga[b]

[a]*Department of Psychology, University of Cincinnati, 2624 Clifton Avenue, Dyer Hall Room 429, Cincinnati, OH 45221, USA*
[b]*Department of Psychology, American University, 4400 Massachusetts Avenue, NW, Washington, DC 20016, USA*

Abstract

Few studies have addressed the amount of training needed to obtain reliable ratings in behavioral observation data. The current study examined the effects of differing intensities of frame-of-reference (FOR) rater training on observers' ratings of anxiety, social skills, and alcohol-specific coping skills in community volunteers with and without social anxiety and alcohol use disorders. Interrater reliability was assessed by comparing three training conditions (no-training, moderate FOR, and intensive FOR) on *discrepancies between raters' scores* and the *strength of association between raters' scoring patterns*. The discrepancies between raters' scores were significantly larger in the control condition than in the intensive and moderate training conditions. Generally, small and nonsignificant differences were found between intensive and moderate training's discrepancy scores. Strength of association results showed significantly lower correlations in the control group compared to the intensive group. However, these correlational results showed less consistent differences between the moderate and other training conditions; differences when found were in the expected directions. Study findings suggest that differing training intensities can affect rating scores and that interrater reliability may be meaningfully assessed through multiple methods.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Alcohol; Coping skills; Behavioral assessment; Role play; Social anxiety; Rater training; Interrater reliability

*Corresponding author. Tel.: +1 513 556 2435; fax: +1 513 556 1904.

E-mail addresses: angkawac@email.uc.edu (A.C. Angkaw), giao.tran@uc.edu (G.Q. Tran), dhaaga@american.edu (D.A.F. Haaga).

Introduction

Rating scales are frequently used as a tool in behavioral assessment to characterize observable impairment in patients with psychological disorders and to evaluate the effects of their treatments. As with any measurement technique, it is critical to obtain reliable ratings from observers because reliability is the foundation of other psychometric properties. While training of observers has been emphasized in behavioral assessment, past investigators have not studied how much training is needed to yield reliable behavioral ratings of anxiety and coping skills in individuals with anxiety and alcohol use disorder(s) (AUD). The current study examined the effects of different training intensities on observers' ratings of anxiety, social skills, and alcohol-specific coping skills exhibited in videotaped role-play interactions of community volunteers with diagnosed social anxiety disorder (SAD) and AUD, SAD only, and no current psychiatric disorders.

Background

Because reliable and valid ratings by observers are critical to behavior assessment, investigators have examined several types of rater training programs, including rater error training (RET) and frame-of-reference (FOR) training. RET aims to improve accuracy by decreasing common rater biases and usually involves exercises designed to increase variability in ratings. RET has been shown to reduce psychometric errors such as leniency error (rater's tendency to give good ratings to all individuals) and halo error (rater's failure to discriminate an individual's performance across the different dimensions of competency); however, when used alone, RET has been shown to decrease accuracy in ratings (Bernardin & Pence, 1980). These findings coincide with substantial research suggesting that reducing psychometric errors has little effect on the accuracy of ratings (McIntyre, Smith, & Hassett, 1984). Instead of focusing specifically on errors, FOR training seeks to aid raters in developing a standard or "frame" of rating by describing the job to be evaluated, giving practice and feedback in ratings, and providing behavioral rationales for ratings (Bernardin & Buckley, 1981; McIntyre et al., 1984). Several studies have shown that participants who completed FOR training were significantly more accurate than those who completed no training or other types of training (McIntyre et al., 1984).

While many investigations have contrasted different types of rater training (Bernardin & Pence, 1980; Heneman, 1988; Martin & Bartol, 1986; Murphy & Balzer, 1989; Pulakos, 1984; Woehr & Huffcutt, 1994), limited research has been conducted to study the effects of varying intensities of rater training (Bernardin, 1978; Ivancevich, 1979). Comparing rater training of varying intensities may facilitate the selection of the rating intensity that may be most appropriate in terms of reliability and time efficiency for a research purpose.

Two studies compared the effects of RET intensities on psychometric quality. Bernardin (1978) contrasted comprehensive RET, abbreviated RET, and two no-training conditions on students' ratings of their nonlaboratory instructors' performance in a longitudinal study. Also in a longitudinal study, Ivancevich (1979) studied the performance evaluations of supervisory engineers who were randomly assigned to an intense RET group, a discussion RET group, or a no-training group. Results of both studies indicated that among the three