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

by

Kelsey Byerly

A Thesis Submitted in Partial Fulfillment of the Degree of Master of Arts in Industrial-Organizational Psychology

Middle Tennessee State University
May 2021

Thesis Committee:

Dr. Mark Frame, Advisor

Dr. Michael Hein, Committee Member

Dr. Alexander Jackson, Critical Reader

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Mark Frame, for his unwavering support and encouragement throughout the completion of this thesis. Your mentorship has helped me to become a capable researcher, and it has postivley impacted my perspective on research and the field of I-O psychology. I scinerely appreciate everything you have done for me over the past two years. I would also like to thank my committee member, Dr. Michael Hein, for your insightfulness and care in ensuring my thesis was done well. Thank you to my critical reader, Dr. Alexander Jackson, for your attnention to detail and inquisitiveness from the start.

I would also like to thank my program mentor, Mary Iseral, for all the resources and support she provided me with in the completion of this thesis. Your guidance is deeply appreciated. Thank you to my family for showing me constant support and helping me achieve such a monumental goal. Lastly, I would like to thank Sarah Tucker and Risa Witherow for always listening and offering advice. Your constant support and friendship means more than you will ever know.

ABSTRACT

This research study sought to determine whether the number of behavioral anchors/examples on a behaviorally anchored rating scaled increased rater accuracy in performance evaluations. Specifically, it examined whether a behavioral anchor rating scale (BARS) with five behavioral anchors/examples yielded more accurate performance evaluation ratings compared to BARS with three behavioral anchors/examples when used to rate the performance of different people performing different teaching assignments. The accuracy of the ratings was assessed in terms of scale alphas, absolute differences between ratings and true score estimates, and over- and under-ratings. In addition, this study explored whether rating accuracy varied among raters with different degrees of conscientiousness. All participants were recruited from Middle Tennessee State University and Amazon's MTurk. Participants were randomly assigned to evaluate the performance of six (6) University professors using either BARS with five behavioral anchors/examples or BARS with three behavioral anchors/examples. Overall, the results demonstrated no significant differences in accuracy between BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples. Additionally, accuracy did not significantly differ among raters with varying degrees of conscientiousness.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I: INTRODUCTION AND LITERATURE REVIEW

Ideally, performance evaluations assess how well an individual is performing on the job using a rating format that provides equitable outcomes for all ratees. Organizations can use this information as a human resource management tool to make decisions such as selection, training, motivation, and compensation. While performance management is a critical tool, organizations often select inadequate rating formats without taking employee performance into consideration. Research has found advantages for some rating forms over others based on the characteristics of the measurement tool (Campbell et al., 1973; Burnaska & Hollmann, 1974; Dickinson & Zellinger, 1980; Schwab et al., 1975). One standardized method that has yielded accuracy and acceptance among employees is behaviorally anchored rating scales (Burnaska & Hollmann, 1974; Campbell et al., 1973; Dickinson & Zellinger, 1980; Schwab et al., 1975).

Behaviorally Anchor Rating Scales

Behaviorally anchored ratings scales (BARS) are evaluation instruments that attempt to capture performance in multidimensional, behavior-specific terms, resulting in an appraisal of performance (Schwab et al., 1975). In other words, BARS consists of items with response options that are exemplars or examples of possible behaviors that might be observed. The use of behavioral anchors provides a clear definition and consistent frame of reference for all behavioral dimensions (DeNisi & Murphy, 2017). This rating format was initially designed by Smith and Kendall (1963) to enhance and standardize observations and thus prepare raters for summary ratings that may be used for future appraisals (Bernadin & Smith, 1981). Smith and Kendall (1963) sought to mitigate

discrepancies in raters' perception of similar ratee behavior, with the goal of establishing valid descriptors of effective and ineffective job behaviors prior to assessing ratee behavior (Bernadin & Smith, 1981).

The BARS format often consists of a vertical graphic rating scale anchored with behavioral examples that include the name, definition, and value of each performance dimension (Dickinson & Zellinger, 1980). The behavioral examples are often derived from observed behaviors through the collection of critical incidents (Flanagan, 1954). Flanagan (1954) defined incidents as behaviors that are either extremely effective or ineffective based on performance expectations. Moreover, these incidents provide an operational definition of the dimension being rated and represent the various degrees of each dimension ranging from effective to ineffective performance (Smith & Kendall, 1963). The overall format of BARS can serve as criteria to evaluate predictors for performance appraisal and review systems (Campbell et al., 1973).

While a few variations are present throughout the literature, Smith and Kendall (1963) proposed a five-phase development plan for BARS. First, critical incidents must be collected in order to provide concrete illustrations of behavior (Flanagan, 1954). Then, the incidents need to be grouped into a smaller set of performance dimensions that are identified and defined by the developers of the rating instrument (Schwab et al., 1975). Next, a panel of subject matter experts (SMEs) are instructed to retranslate, or reallocate, the critical incidents into the performance dimensions (Schwab et al., 1975). This step requires SMEs to assign each incident to the dimension it best describes. A retranslation criterion is established as a basis for determining which incidents will be retained for the final instrument. Once the incidents are retranslated, the same group of experts must rate

the behavior described in the incident based on a scale ranging from effective to ineffective, with the number of scale points and behavioral anchors/examples specified by the developer of the instrument. For the purpose of this thesis, a comparison in terms of rater accuracy will be made between two five-point BARS where one will have three behavioral anchors/examples and another will have five behavioral anchors/examples. The average rating for each incident indicates the degree to which it is representative of effective performance on a particular dimension (Schwab et al., 1975). The amount of agreement among raters is expressed in terms of standard deviation (Schwab et al., 1975). BARS developers compute the mean effectiveness values for each incident and use the standard deviation of the SMEs' ratings to determine the degree of agreement. The lower the standard deviation, the greater the agreement among raters. Typically, incidents that have standard deviations of 1.50 or less are retained when using a 7-point scale (Schwab et al., 1975). This threshold is to be adjusted when taking into account the number of scale points on a BARS. Lastly, a subset of incidents that meet the retranslation criteria and standard deviation criteria are used as behavioral anchors for the performance dimensions.

The major advantage of BARS over alternative rating formats is credited to the rigor of the development process of the scale. As noted by the five-step plan previously outlined, BARS involves the participation of individuals who are familiar with the job and adds value to generating incidents that cover the performance domain, resulting in a content valid instrument (Schwab et al., 1975). The retranslation step in BARS is designed to eliminate incidents that do not clearly align with a single dimension. This step also helps to identify unclear or poorly defined dimensions (Smith & Kendall, 1963).

In addition, the incidents and dimensions generated by incumbents is likely to result in terminology that is meaningful and unambiguous to the individuals who will ultimately use the instrument (Schwab et al., 1975). Smith and Kendall (1963) suggest that retranslation and retention of incidents with adequate agreement among raters may reduce errors such as leniency and central tendency.

Another advantage of BARS pertains to psychometric properties. In comparison to summated rating scales, BARS produces less leniency error, less halo error, and less method variance (Campbell et al.,1973). This finding was further supported by Burnaska and Hollmann (1974) who indicated that BARS reduce the amount of leniency error and increase the amount of variance attributed to ratee performance. However, the researchers found no evidence for any one superior rating format. Dickinson and Zellinger (1980) suggest that BARS and mixed standard scales format are both equally desirable in their psychometric properties when both formats use behavioral examples. Although, the BARS format was found to be the preferred rating instrument because it provides an opportunity for feedback and goal setting (Dickinson & Zellinger, 1980).

Despite the appealing characteristics of BARS, research has noted drawbacks of the five-step procedure used to create the final instrument. Specifically, critical incidents collected in step one are often lost at the retranslation step due to the standard deviation criterion (Schwab et al., 1975). The loss of incidents makes it difficult for raters to assign the remaining incidents to a behavioral dimension as well as determine the degree of effectiveness of the observed behavior relative to other examples (Schwab et al., 1975). While altering the set criterion appears to be a plausible solution, Schwab and colleagues (1975) indicate some potentially negative implications for the construct validity of the

final scale. For instance, less stringent retranslation criteria may result in ambiguity in the final instrument (Schwab et al., 1975). Therefore, scale developers should aim to retain behavioral examples that are clearly distinct from one another.

Hauenstein and colleagues (2010) discuss the issue of missing middle anchors. The researchers identified two general reasons for the missing middle anchor problem: (1) job incumbents tend to provide more extreme examples of behavior than average behaviors due to the nature of the critical incident technique, and (2) the agreement among raters tends to be lower for examples of average performance. The latter issue often results in a greater chance that the average behavioral examples fail to meet the standard deviation criterion, therefore those incidents never make it past the retranslation phase. Based on this evidence, it is clear that BARS development should encourage raters to identify as many average behaviors as possible. The lack of average behaviors can be resolved through training by emphasizing the meaning of the performance dimensions and practicing frame-of-reference to increase interrater agreement (Hauenstein et al., 2010).

The literature suggests that minimizing rating errors does not necessarily maximize rating accuracy. Therefore, different development procedures and formats may be required in order to further validate BARS and assess the accuracy of BARS. One possible development and format issue that might impact accuracy is the number of behavioral anchors included on a BARS. The number of anchors may increase or decrease the accuracy of BARS instruments, and consequently, provide more effective performance feedback. Unfortunately, research has yet to explore in-depth the intended effects of including more or fewer anchors using BARS format with regard to

psychometric properties and overall accuracy. Lissitz and Green (1975) found the number of scale points, such as a five point rating scale versus a two point rating scale, is accompanied by an increase in reliability. However, the optimal degree of reliability levels off after 5 scale points (Lissitz & Green, 1975). Iseral (2020) further investigated this idea in a recent study that evaluated employment interviews using BARS with three behavioral anchors/examples versus five behavioral anchors/examples. The findings suggest that only the over- and under-composite ratings improved when using a BARS with five behavioral anchors/examples (Iseral, 2020). In other words, BARS with five behavioral anchors/examples yielded less rater error in terms of leniency and severity. However, the evidence was sufficient enough to conclude that five behavioral anchors/examples yield more accurate ratings in comparison to three behavioral anchors/examples. While this study analyzed the use of BARS for rating employment interviews, it is unknown whether these results will translate to similar findings when evaluating performance appraisals using BARS.

The overarching goal of BARS is to create a record of behavioral incidents that clearly outline expected performance and ultimately provide a summary rating format that can be used for performance evaluations (Bernadin & Smith, 1981). This rating format is aimed to define, clarify, and operationalize the observations of the rater (Bernardin & Smith, 1981). Improving the scaling process for BARS is imperative because BARS remains a popular scale format for practitioners and researchers.

Performance Evaluations

Performance evaluations serve as valuable human resource management tools that help make decisions such as selection, training, motivation, and compensation.

Performance evaluations also serve as a resource for effective coaching and development of personnel. The behaviors evaluated during appraisal interviews have critical implications for both the individual and the organization, so accuracy in measuring performance is of major concern. It is important to keep in mind that performance appraisal systems are embedded in organizational contexts. Understanding the nature of an organization will facilitate the development and implementation of an effective performance appraisal system (Wilson et al., 2017).

Despite their prevalence, many performance appraisal systems are not appropriately developed or implemented. Ratings are widely viewed as inaccurate and biased, especially if raters are not trained to identify their own biases or errors (Wilson et al., 2017). In many organizations, ratings are highly inflated for multiple reasons, such as discomfort with the appraisal system or concern with maintaining productive work relations (Wilson et al., 2017). Employees who receive appraisals may perceive the process as unfair and may not always act on feedback provided by leaders. According to Latham et al. (2005) employee performance is just as likely to decrease as it is to increase after receiving appraisal feedback. A successful performance appraisal system requires a deep understanding of the performance domain, the critical tasks, and work behaviors that reflect an individual's contributions to an organization. To achieve the most effective results, organizations and leaders should view performance appraisals as an integral part of an ongoing performance management system designed to motivate employees to improve performance (Wilson et al., 2017). In addition, performance appraisals should not be viewed as an isolated human resources function, but rather part of an integrated

human resources system that includes selection, training, compensation, and any other human resource practice.

Although performance appraisals are widely used and serve various human resource functions, they are widely disliked because of problems associated with their utility and accuracy. DeNisi and Pritchard (2006) indicate that relatively few employees believe their organization's performance appraisal helps them improve performance. The most frequently cited problems with appraisal systems concern both the rater and the rating instrument. For instance, raters are too lenient, raters fail to differentiate between workers, and the most critical behaviors are not being evaluated (Wilson et al., 2017). Rating dimensions may not capture important aspects of the job, and performance appraisal ratings are often not linked to organizational decisions, such as who to promote or terminate (Wilson et al., 2017). Performance appraisals also force raters to make distinctions among employees, which can discourage employees to make changes to their performance (Wilson et al., 2017).

While there is no doubt that there are many issues related to performance appraisals, if proper attention is given to the construction of performance appraisal systems and administration of evaluations, appraisals can be a useful and valued resource for organizations. Organizations often view performance appraisals as a menial task (Wilson et al., 2017). However, if performance appraisals were perceived in a broader context, then organizations would recognize them as a tool for improving performance (Wilson et al., 2017). BARS has demonstrated a psychometrically sound rating scale format that uses behavioral examples to enhance the performance of individuals.

Performance Evaluations Using Behavioral Items

Borman (1979) suggested that raters tend to experience trouble with the specificity of behavioral anchors. In response to this challenge, Borman (1979) proposed anchoring the scale with more general behavioral incidents in order to help raters match observed behaviors with the scale. For instance, Smith and Kendall (1963) provided more abstract anchors for each behavioral incident along each dimension, which was believed to help raters understand how to assess a ratee's performance using behavioral anchors. Bernardin and Smith (1981) found that the inclusion of general behavioral examples significantly improved the quality of ratings. By making this adjustment, the behavioral dimensions are made clearer to the raters, and the scores given to ratees are more likely to accurately reflect their performance (Borman, 1979).

One criticism of this approach is the time involved to write and scale incidents for each dimension (Bernardin & Smith, 1981). Although this is a reasonable concern, Bernardin & Smith (1981) argue that the time required to write and scale incidents on BARS is well worth it, given the need to document and justify ratings during performance appraisal interviews. In addition, the time spent recording behavioral observations for each ratee may not be much greater than the time required to rate the same number of ratees on other scales such as a summated scale (Bernardin & Smith, 1981). It is evident from the literature that BARS cannot be developed without a record of observed behaviors or a clear understanding of expected performance. Smith and Kendall (1963) stress that observed behaviors must be recorded to maintain the integrity of the anchors included in the final instrument. This procedure is important for enhancing

future observations and standardizing the observational process (Bernardin & Smith, 1981).

Performance Evaluations and Conscientiousness

Raters differ in how they evaluate ratees based on performance dimensions. This difference may be a function of personality (Bernardin et al., 2000; Ogunfowora et al., 2010; Tznier et al., 2002). Conscientiousness, a facet of personality, is related to overall performance in a wide range of jobs, and it is likely to influence how individuals carry out their job tasks and responsibilities, such as conducting performance appraisals (Tziner et al., 2002).

Costa and McCrae (1992) suggest that individuals high in conscientiousness have high performance standards and strive for excellence. Therefore, raters with a high level of conscientiousness might feel a greater responsibility to carry out performance evaluations with diligence, resulting in better discrimination within performance dimensions and between ratees (Tziner et al., 2002). Ogunfowora and colleagues (2010) suggest that conscientious raters tend to be careful and diligent in their evaluations. Conscientious raters are likely to make an effort to provide accurate evaluations, and they are more prone to internalize the rules and procedures that define the appraisal system (Tziner et al., 2002). In addition, conscientious raters may be less swayed by their own attitudes, beliefs, and biases when evaluating ratees (Tziner et al., 2002). Raters who demonstrate conscientious characteristics will be less inclined to rate on the basis of their attitudes toward the performance appraisal system or organization. Rather, the ratings will reflect the ratee's true performance (Tziner et al., 2002). These assumptions made by

Tziner and colleagues (2002) suggest that conscientious raters should provide more accurate ratings than individuals with lower levels of conscientiousness.

Ogunfowora et al. (2010) found a negative relationship between rater conscientiousness and overall performance ratings. In other words, individuals with high levels of conscientiousness provided lower overall performance evaluation ratings (Ogunfowora et al., 2010). Conscientious raters tend to be careful and diligent in their evaluations, and thus they are less likely to allow beliefs and biases to influence their final ratee evaluations (Ogunfowora et al., 2010). Bernardin and colleagues (2000) also determined that conscientious raters tend to provide lower overall mean ratings. Additional research expands upon this finding by including the interaction of identifiability, or accountability. Roch et al. (2005) showed that raters who were high on conscientiousness provided relatively lower ratings regardless of the level of accountability. However, raters who were low on conscientiousness tended to inflate their ratings when identified (Roch et al., 2005). In addition, Roch and colleagues (2005) found that highly conscientious individuals reported feeling significantly more accountable than less conscientious individuals. It appears the relationship between conscientiousness and performance ratings is more complex than the main effect proposed by Bernardin and colleagues (2000).

Measuring Performance Evaluations

Performance ratings are often used as the only criteria for validating personnel decisions such as selection or promotion. Unfortunately, performance ratings are almost always contaminated by various forms of rater errors. The effort to minimize rater errors has centered around three areas. Borman (1978) suggests that the accuracy of ratings

often depends on the rater's opportunity to observe relevant, performance-related behaviors. In a prior research study, Borman (1974) found that raters displayed greater interrater reliability when they had the opportunity to observe ratee behaviors and engage in the development of performance dimensions. Second, rater training has proven successful at reducing various rating errors (Borman, 1978). Lastly, Smith and Kendall (1963) constructed a rating scale that operationally defines performance behaviors, which has been shown to sometimes reduce rating error significantly (e.g., Campbell et al., 1973). Although the review on a superior scale format is mixed, a well-developed scale with defined dimensions will yield higher quality ratings than a scale with dimensions that are not operationalized (Borman, 1978).

Furthermore, Borman (1978) classified rater error with performance ratings into four distinct categories. The first issue considers the lack of opportunity to observe job behaviors needed to construct a rating format such as BARS. The Smith and Kendall (1963) approach to BARS requires subject matter experts to identify relevant job behaviors that would help define effective performance. Borman (1978) also indicated that raters are unaware of common rating errors and lack experience in making performance evaluations, both of which contribute to high-quality ratings. However, training has been shown to help raters understand the rating process and teach raters how to differentiate performance dimensions using the anchors provided. Third, certain rating formats prevent raters from matching observed job-related behaviors to a specific level or anchor on the performance scale. This may be attributed to the selection of rating formats that are not user-friendly or accepted by raters. BARS combat this issue by introducing concise behavioral statements associated with scale points. Lastly, Borman (1978)

mentioned that organizational factors often force performance evaluations to reflect organizational demands rather than consider employee performance. For instance, raters often feel pressured to provide high ratings when they are required to discuss the performance ratings with the ratee, which forces performance evaluations to reflect organizational demands rather than the true levels of performance exhibited by ratees (Borman, 1978).

Although BARS are a useful tool for evaluating the performance of individuals, the ratings are often impacted by rater error which decreases the accuracy of the ratings. Therefore, it is important to ensure the raters are aware of common psychometric errors in order to generate accurate ratings. Otherwise, the performance appraisal system will not improve individual-level performance, and consequently, lead to negative organizational level implications.

Accuracy

Rating errors arise because raters might not have the opportunity to observe job-relevant behaviors, or they rely on stereotypes of exceptional and unexceptional employees. In addition, raters have a tendency to evaluate individuals who are similar to themselves more highly than those who are not, making the rating process less objective. Some errors are intentional and other errors are unintentional depending on the intended purpose of the ratings. These errors can be detected by looking at the performance appraisal data. The utility of a rating scale, such as BARS, is only as good as the accuracy of the ratings. Sulsky and Balzer (1988) describe the accuracy of measurement as the "strength and kind of relation between one set of measures and a corresponding set of measures considered to be an accepted standard for comparison" (p. 497). In other words,

the accuracy of ratings is computed by comparing the rater's performance evaluations for ratees with corresponding evaluations provided by expert raters (Sulsky & Balzer, 1988). However, a variety of different methods exist for comparing two sets of ratings. The literature surrounding accuracy, primarily rater error, focuses on three psychometric issues: the susceptibility of BARS to leniency effects, whether or not BARS get at independent performance dimensions, and the reliability of BARS.

Leniency error is evident when a rater uses only the positive end of the rating scale and thus fails to differentiate among ratees when evaluating performance (Kingstrom & Bass, 1981). A similar outcome occurs when a rater uses only the negative end of the rating scaled, which is referred to as severity error (Sulsky & Balzer, 1988). In general, research has determined a higher mean rating to be indicative of higher susceptibility to leniency effects. Smith and Kendall (1963) presumed that BARS is less susceptible to leniency effects because the development procedure yields clearly defined dimensions and anchors. To corroborate this hypothesis, Campbell et al. (1973) found that BARS produced less leniency error in comparison to a summated format. In addition, Murphy and colleagues (1982) found a relationship between accuracy in observation and accuracy in performance evaluation. The results indicate that raters who overestimate the frequency of favorable behavior also tend to give higher performance ratings (Murphy et al., 1982).

Halo effect is indicated by the influence of a single dimension of an individual's performance on other dimensions, resulting in consistently high, medium, or low ratings on all performance dimensions (Lunenburg, 2012). This measure of accuracy is typically assessed by comparing correlations or variances in ratings across dimensions (Sulsky &

Balzer, 1988). Thus, halo examines the ability to generate independent dimensions that are multidimensional measures of performance (Schwab et al., 1975). The lower the intercorrelations, the less halo error or similarity across performance dimensions. When independence is not achieved, the scale is considered redundant and thus fails to provide unique performance information (Schwab et al., 1975). Campbell et al. (1973) found that dimensions on their BARS format demonstrated somewhat lower intercorrelations than a numerically scaled instrument. However, the comparisons between correlations do not consider individual ratings and corresponding true scores, so halo does not provide a direct estimate of rating accuracy (Sulsky & Balzer, 1988). As a result, halo should not be considered in isolation when determining the accuracy of ratings.

The reliability of BARS is often evaluated from the critical incidents and retranslation steps by correlating average scale values assigned to the anchors by both groups (Smith & Kendall, 1963). Schwab et al. (1975) noted that high scale reliabilities may occur when similar groups of raters take part in the development of BARS. However, high scale reliabilities can be achieved without generating an operationally effective scale by reducing the number of behavioral anchors per dimension (Borman, 1974). In contrast, Borman (1974) suggested that low reliabilities may be a result of raters at different levels within the organization who may have different ratee performance expectations that they feel are necessary to evaluate. From an applied perspective, the ability of different raters to agree on an adequate assessment of ratees is more important than scale reliabilities (Schwab et al., 1975).

In spite of these rating errors, the literature demonstrates the success of some training programs at mitigating certain rating errors. Historically, lectures are the most

traditional form of rater training, which educates raters on common errors and caution them to avoid such errors. This type of training program initially appeared to be effective because it led to lower levels of intercorrelations and lower mean ratings, which was thought to indicate less leniency and halo error (DeNisi & Murphy, 2017). Later research determined this logic to be incorrect, which led to various modifications of rater training. Borman (1979) pioneered a training approach that has become a prominent form of training, Frame of Reference (FOR) training. This approach involves showing the raters videotapes of performance, asking them to rate the tapes, and then discussing the most appropriate ratings for each candidate. In terms of practical implications, an organization should conduct a cost-benefit analysis regarding the training of raters to reduce various types of error. While the literature indicates that training has been effective in reducing halo and leniency (Borman, 1979), further exploration of training as it relates to the number of behavioral anchors and accuracy is unfortunately outside the scope of this thesis.

Behaviorally Anchored Rating Scales and Accuracy

When using rating scales, it is essential for raters to have an understanding of what is meant by accuracy and how to achieve accurate ratings. Accuracy describes both the strength and kind of relationship between one set of measures and a corresponding set of measures (Guion, 1995). The accuracy of ratings has been measured in a number of ways. For instance, some methods involve a comparison of ratings to true score measures of performance, while others use differential accuracy, which requires raters to make a differentiation among ratees within each performance dimension (Sulsky & Balzer, 1988). The direct measurement of rating accuracy requires a standard against which

ratings can be evaluated. This standard is often referred to as a true score and is thought to represent a rating that is provided by an unbiased, careful rater who completed ratings under optimal conditions (Murphy & Cleveland, 1995). Thus, the validity of an accuracy measure depends on true scores that are used to evaluate the ratings. Murphy and Cleveland (1995) define true scores as the expected value obtained from a specific population in which the raters are considered the population. True scores can be obtained using various procedures. One method uses the average scores provided by all subjects or raters as the true score (Sulsky & Balzer, 1988). Another procedure for developing true scores involves averaging the ratings of critical incidents used as information in ratees' performance profile (Sulsky & Balzer, 1988). The most widely used strategy for estimating true scores was developed by Borman (1977). This procedure involves computing average scores among expert raters, which yields true score ratings for each ratee on every performance dimension Borman (1977). The individual rater scores are then compared with the true scores. The closer the rater's ratings are to the true score ratings, the more accurate the ratings are perceived to be. (Borman, 1977).

Rater error measures are used to infer accuracy indirectly. The presence of rater errors generally indicates inaccurate ratings, and the absence of rater errors suggests accuracy (Murphy & Cleveland, 1995). Accuracy scores are often evaluated against three sources of error: leniency, halo, and reliability. Leniency and severity, otherwise known as over- and under-rating, are evident when a rater tends to use the positive or negative half of the rating scale to the relative exclusion of the opposite end (Schwab et al., 1975). Halo and negative halo occur when a rater tends to endorse the extreme ends of the rating scale across multiple dimensions or behaviors due to their perception that the person they

are rating excels or is poor in some areas and thus should be rated high or low in all domains of performance (Schwab et al., 1975). As a result of these errors, the ratings provided fail to differentiate among ratees when evaluating performance (Kingstrom & Bass, 1981). One way to measure leniency is by calculating the positive or negative difference between the rater's rating and the true score rating (Sulsky & Balzer, 1988). Lastly, the reliability of a scale refers to the consistency achieved when ratings are provided for each ratee across the performance dimensions (Harrelll & Wright, 1990). The more consistent the results, the higher the reliability of the scale; the less consistent the results, the lower the reliability of the scale (Harrell & Wright, 1990). Further, a scale must be reliable if it is to provide accurate results (Harrell & Wright, 1990).

Iseral (2020) found that using a BARS with five versus three anchors/examples yielded greater accuracy for employment interviews. If these findings hold true for performance appraisals, there is potential for five anchor/example BARS to become an even more accurate rating format than the three anchor/example BARS. If five anchor/example BARS are more accurate than three anchor/example BARS, organizations should consider using this type of rating format to conduct performance appraisals. This leads to the following hypotheses:

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

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

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

Hypothesis 1c: BARS with five anchors/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 anchors/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 assessed with BARS with three anchors/examples compared to BARS with five anchors/examples?

RQ 1b: Are there differences in the scale alphas among low conscientious raters on performance evaluation ratings assessed with BARS with three anchors/examples compared to BARS with five anchors/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 rating) vary among high conscientious raters on performance evaluation ratings assessed with BARS with three anchors/examples compared to BARS with five anchors/examples?

RQ 2b: Do accuracy estimates (absolute value and over-under rating) vary among low conscientious raters on performance evaluation ratings assessed with BARS with three anchors/examples compared to BARS with five anchors/examples?

CHAPTER II: METHOD

The study was a between- and within-subjects design to assess the accuracy of performance evaluation ratings on a computerized assessment. The study used videos of six (6) University professors giving lectures on course topics at the undergraduate level as stimulus material for the participants. Participants were randomly assigned to provide ratings using either BARS with three anchors/examples or BARS with five anchors/examples. The research was conducted in accordance with the Middle Tennessee State University Institutional Review Board.

Participants

Some participants were recruited using MTSU's SONA participant pool. SONA is an online research management system that allows students to serve as a participant in a study. By recruiting from SONA, the researchers were able to collect information from undergraduate students with diverse backgrounds. The students received course credit in exchange for their participation. In addition to participant recruitment at MTSU, the researchers recruited participants using Amazon's MTurk, an online crowdsourcing platform that allows researchers to obtain data from a large pool of individuals. These participants received three dollars in compensation for their participation in the study.

Approximately 204 participants began the study. However, only the participants who passed the attention check process were included in this analysis. To pass the attention check process, participants had to answer 12 out of 36 attention check questions correctly. Attention check questions required participants to correctly identify the gender of the professor and subject matter of each lecture. The final number of participants included 77 (41 men, 35 women, and 1 that preferred not to report gender). The ages

ranged from 18 to 63. The most frequently reported ethnicity was Caucasian/White (62.3%, followed by African-American/Black (19.5%), Asian/Pacific Islander (3.9%), those that preferred not to say (3.9%), Arab/Middle Eastern (2.6%), and Hispanic/Latino (2.6%), Bi-Racial/Multi-Racial (2.6%), Native American/Alaskan Native (1.3%), and those who indicated other (1.3%). Of those who indicated they are currently enrolled at a college or university (77%), the most frequently reported level in college was Freshman (22.1%), Sophomore (13%), Junior (6.5%), Senior (3.9%), and Graduate Student (10.4%). The 77 participants were randomly assigned to one of two BARS conditions with 41 participants in the five behavioral anchors/examples condition and 36 participants in the three behavioral anchors/examples condition.

Measures

Participant Information Questionnaire

Participants were asked to respond to questions at the beginning of the survey to assess participants' knowledge and experience with performance evaluations. They were asked to indicate approximately how many teacher performance evaluations they had completed within the last year, and they were presented with questions to assess their attitudes toward performance evaluations. Participants were asked to indicate if they were familiar with behaviorally anchored rating scales.

Course Lecture Videos

The course lecture videos were obtained from a performance evaluation training project facilitated by James Parker, an MTSU graduate student. Six course lecture videos were selected from the training project, and each video covered a topic in the following undergraduate level courses: Introduction to Psychology, Introduction to Philosophy,

Introduction to World History, Developmental Psychology, Computer Science, and Physics. The course lecture videos were approximately one and a half minutes to three and a half minutes in length. Each participant was asked to watch the same six course lecture videos regardless of condition.

Behaviorally Anchored Rating Scales

Participants watched all six videos of University professors lecturing on various undergraduate course topics, and then they were asked to rate the professors' performance accordingly.

James Parker developed five performance dimensions with language and formatting that matched the lecture videos used as a performance evaluation tool approved by the MTSU Department of Psychology. The content of the performance dimensions was further revised for this study to align with the content in the course lecture videos, resulting in four performance dimensions. The four performance dimensions included presentation ability, organization and clarity, intellectual and scholarly, and incorporating student interaction. Please see Appendix A for the definitions pertaining to each dimension.

Each performance dimension varied in the number of performance evaluation ratings provided. Presentation ability consisted of four ratings, organization and clarity included four ratings, intellectual and scholarly comprised of two ratings, and incorporating student interaction involved three ratings, resulting in a total of 13 performance evaluation ratings. Participants were asked to rate all 6 professors using the same 13 performance evaluation ratings with the only difference being the number of behavioral anchors/examples provided.

The BARS created by James Parker initially consisted of three behavioral anchors/examples for each dimension and performance evaluation rating. Two additional behavioral anchors/examples were created by the researchers for each performance evaluation rating to make a BARS with five behavioral anchors/examples.

BARS with Five Anchors.

Each performance evaluation rating was presented on a five-point Likert scale with five behavioral anchors/examples of performance. Please see Appendix B for the BARS with five behavioral anchors/examples.

BARS with Three Anchors.

Each performance evaluation ratings were presented on a five-point Likert scale with three behavioral anchors/examples of performance. Please see Appendix C for the BARS with three behavioral anchors/examples.

Overall Performance Rating

Regardless of condition, participants were asked to provide an overall performance rating for each professor which was independent of the BARS with five and three behavioral anchors/examples. The overall performance rating scale consisted of a five-point Likert scale (1 = Awful; 5 = Excellent).

Attention Check

For each course lecture video, we included attention check questions to ensure participants were paying attention to the University professor's performance during the course lecture. Participants were asked to indicate a response to the following six attention check questions: the professor's gender, course subject, the topic discussed, if the professor walked around or sat on the desk, and whether or not the professor was

wearing glasses. For instance, participants were asked, “Which topic was discussed during the lecture,” and, “Did the professor sit on the desk during the lecture?” Each participant, regardless of the condition, received the same six attention check questions per each course lecture video. As noted earlier, only the professor gender and course subject attention check items were used for the present study. Please see Appendix D for the full list of all attention check questions.

Manipulation Check

In both conditions, after viewing the videos and providing performance evaluation ratings for each professor, participants were asked a few questions to determine if they attended to the manipulation of the BARS. Specifically, participants were asked to indicate how many behavioral anchors/examples they were presented with and what side of the screen had behavioral anchors/examples of good performance versus bad performance. Each participant, regardless of the condition, received both manipulation check questions. Please see Appendix E for the manipulation check questions.

HEXACO

After evaluating the performance of University professors, participants were asked to complete a short personality inventory by responding to items based on the HEXACO personality model. The measure assessed candidates on six dimensions: honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience. Participants responded to an index consisting of 60 items assessed along a 5-point Likert scale (1 = Strongly Disagree; 5 = Strongly Agree).

According to the HEXACO personality model, an individual who scores high on the conscientiousness scale organizes their time, strives for accuracy, and carefully

deliberates decisions. In contrast, an individual with a very low score on conscientiousness is characterized as impulsive and satisfied with work that contains a few errors. Please see Appendix F for the HEXACO.

Demographic Questionnaire

At the end of the study, participants were asked to respond to demographic questions. These questions asked participants to indicate their age, year in college, current major, gender, and ethnicity. Participants were provided with a debriefing statement and asked not to discuss the study with their peers or others that may have participated in the study.

Procedure

Prior to conducting the study, the researchers identified videos of University professors giving lectures to their class which served as the stimulus material. The videos involved lectures on topics within the following courses: Introduction to Psychology, Introduction to Philosophy, Introduction to World History, Developmental Psychology, and Computer Science, and Physics. After compiling the videos, the researchers developed BARS with five behavioral anchors/examples that evaluate the performance dimensions as previously discussed in the measures section. The researchers adapted the BARS with five behavioral anchors/examples to create BARS with three behavioral anchors/examples by removing the second and fourth anchor/example while retaining a five-point Likert scale.

The participants were recruited from MTSU's SONA participant pool and Amazon's MTurk. First, they were asked to indicate how many teacher evaluations they had completed within the last year, their attitudes toward performance evaluations, and

their familiarity with BARS. Then, the online survey platform randomly assigned participants to one of the two conditions: BARS with five behavioral anchors/examples or BARS with three behavioral anchors/examples. Participants in each condition viewed a set of instructions regarding the purpose of BARS and how BARS are used.

Each participant watched the same six course lecture videos regardless of condition. After watching each video, participants were asked to rate the performance of the professor using either BARS with five behavioral anchors/examples or three behavioral anchors/examples. Then, they were asked to rate the professor's overall performance on a five-point Likert scale (1 = Awful; 5 = Excellent), which was independent of the BARS with five behavioral anchors/examples or three behavioral anchors/examples. In addition, participants were given attention check questions throughout the study to ensure they were paying attention to the lecture. After providing ratings for each professor, participants responded to two manipulation check questions. Next, they were asked to complete a personality inventory using items from the HEXACO. Lastly, participants were asked to respond to demographic questions. When participants completed the study, they were awarded course credit or monetary payment for their participation.

Expert Ratings

True score estimates provide the basis for several evaluation criteria used in this study. The videos were viewed and rated by four graduate students to obtain true score estimates for each course lecture video. The true score estimates were obtained in a manner suggested by Sulsky and Balzer (1988). Specifically, expert raters watched all six course lecture videos and evaluated the performance of the professors using BARS with

five behavioral anchors/examples. The expert raters were asked to calibrate their ratings by reaching a consensus regarding the level of performance of each professor (Sulsky & Balzer, 1988). The expert ratings were produced under optimal conditions (Murphy & Cleveland, 1995). The expert raters were allotted as much time as needed to view the course lecture videos and provide ratings.

The four expert raters were individuals who were experienced with performance evaluations and behaviorally anchored rating scales. The expert ratings can be seen in Tables 1 through 6. The participants' ratings were compared to the expert ratings in order to obtain accuracy scores. In addition, the absolute differences between ratings were compared. Reliability estimates were computed for the BARS with five behavioral anchors/examples and the BARS with three behavioral anchors/examples. This allowed the researchers to not only compare participants to experts, but also to one another.

Table 1

Expert Ratings for Introduction to Psychology

Presentation Ability	Ratings
Please rate the professor's ability to show excitement about the course.	5.00
Please rate the professor's ability to use presentation aids.	1.00
Please rate the professor's ability to communicate information about the course.	4.00
Please rate the professor's ability to communicate information with or without the use filler words.	5.00
Organization and Clarity	
Please rate the professor's ability to present the course material in an organized, logical framework.	4.00
Please rate the professor's ability to teach at a pace that supports learning.	5.00
Please rate the professor's ability to teach at a level that supports learning.	4.00
Please rate the professor's ability to make an effort to gauge students' understanding of the course material.	4.00
Scholarly and Intellectual	
Please rate the professor's ability to focus on relevant course material.	4.00
Please rate the professor's ability to draw upon scholarly works.	1.00
Incorporating Student Interaction	
Please rate the professor's ability to provide students with an opportunity to participate.	5.00
Please rate the professor's ability to ask questions that deepened students' understanding of the course material.	5.00
Please rate the professor's ability to allow students to provide meaningful contributions.	5.00
Overall Rating	
Please rate the professor's overall performance.	4.75

Table 2

Expert Ratings for Introduction to Philosophy

Presentation Ability	Ratings
Please rate the professor's ability to show excitement about the course.	5.00
Please rate the professor's ability to use presentation aids.	2.00
Please rate the professor’s ability to communicate information about the course.	4.25
Please rate the professor’s ability to communicate information with or without the use filler words.	5.00
Organization and Clarity	
Please rate the professor's ability to present the course material in an organized, logical framework.	3.00
Please rate the professor’s ability to teach at a pace that supports learning.	2.00
Please rate the professor’s ability to teach at a level that supports learning.	3.00
Please rate the professor's ability to make an effort to gauge students' understanding of the course material.	2.50
Scholarly and Intellectual	
Please rate the professor's ability to focus on relevant course material.	5.00
Please rate the professor's ability to draw upon scholarly works.	1.00
Incorporating Student Interaction	
Please rate the professor's ability to provide students with an opportunity to participate.	3.00
Please rate the professor's ability to ask questions that deepened students' understanding of the course material.	3.00
Please rate the professor's ability to allow students to provide meaningful contributions.	3.00
Overall Rating	
Please rate the professor's overall performance.	4.00

Table 3

Expert Ratings for Introduction to World History

Presentation Ability	Ratings
Please rate the professor's ability to show excitement about the course.	3.00
Please rate the professor's ability to use presentation aids.	1.00
Please rate the professor’s ability to communicate information about the course.	3.00
Please rate the professor’s ability to communicate information with or without the use filler words.	2.00
Organization and Clarity	
Please rate the professor's ability to present the course material in an organized, logical framework.	1.00
Please rate the professor’s ability to teach at a pace that supports learning.	2.00
Please rate the professor’s ability to teach at a level that supports learning.	2.00
Please rate the professor's ability to make an effort to gauge students' understanding of the course material.	1.00
Scholarly and Intellectual	
Please rate the professor's ability to focus on relevant course material.	3.00
Please rate the professor's ability to draw upon scholarly works.	1.00
Incorporating Student Interaction	
Please rate the professor's ability to provide students with an opportunity to participate.	1.00
Please rate the professor's ability to ask questions that deepened students' understanding of the course material.	1.00
Please rate the professor's ability to allow students to provide meaningful contributions.	1.00
Overall Rating	
Please rate the professor's overall performance.	2.00

Table 4

Expert Ratings for Developmental Psychology

Presentation Ability	Ratings
Please rate the professor's ability to show excitement about the course.	4.00
Please rate the professor's ability to use presentation aids.	3.00
Please rate the professor’s ability to communicate information about the course.	5.00
Please rate the professor’s ability to communicate information with or without the use filler words.	4.00
Organization and Clarity	
Please rate the professor's ability to present the course material in an organized, logical framework.	4.00
Please rate the professor’s ability to teach at a pace that supports learning.	3.00
Please rate the professor’s ability to teach at a level that supports learning.	3.00
Please rate the professor's ability to make an effort to gauge students' understanding of the course material.	1.00
Scholarly and Intellectual	
Please rate the professor's ability to focus on relevant course material.	5.00
Please rate the professor's ability to draw upon scholarly works.	4.00
Incorporating Student Interaction	
Please rate the professor's ability to provide students with an opportunity to participate.	1.00
Please rate the professor's ability to ask questions that deepened students' understanding of the course material.	1.00
Please rate the professor's ability to allow students to provide meaningful contributions.	1.00
Overall Rating	
Please rate the professor's overall performance.	4.00

Table 5

Expert Ratings for Computer Science

Presentation Ability	Ratings
Please rate the professor's ability to show excitement about the course.	3.00
Please rate the professor's ability to use presentation aids.	4.00
Please rate the professor’s ability to communicate information about the course.	4.00
Please rate the professor’s ability to communicate information with or without the use filler words.	5.00
Organization and Clarity	
Please rate the professor's ability to present the course material in an organized, logical framework.	3.00
Please rate the professor’s ability to teach at a pace that supports learning.	5.00
Please rate the professor’s ability to teach at a level that supports learning.	5.00
Please rate the professor's ability to make an effort to gauge students' understanding of the course material.	5.00
Scholarly and Intellectual	
Please rate the professor's ability to focus on relevant course material.	5.00
Please rate the professor's ability to draw upon scholarly works.	1.00
Incorporating Student Interaction	
Please rate the professor's ability to provide students with an opportunity to participate.	5.00
Please rate the professor's ability to ask questions that deepened students' understanding of the course material.	4.00
Please rate the professor's ability to allow students to provide meaningful contributions.	4.00
Overall Rating	
Please rate the professor's overall performance.	4.00

Table 6

Expert Ratings for Physics

Presentation Ability	Ratings
Please rate the professor's ability to show excitement about the course.	4.00
Please rate the professor's ability to use presentation aids.	3.00
Please rate the professor’s ability to communicate information about the course.	3.00
Please rate the professor’s ability to communicate information with or without the use filler words.	5.00
Organization and Clarity	
Please rate the professor's ability to present the course material in an organized, logical framework.	1.00
Please rate the professor’s ability to teach at a pace that supports learning.	3.00
Please rate the professor’s ability to teach at a level that supports learning.	3.00
Please rate the professor's ability to make an effort to gauge students' understanding of the course material.	2.00
Scholarly and Intellectual	
Please rate the professor's ability to focus on relevant course material.	4.00
Please rate the professor's ability to draw upon scholarly works.	1.00
Incorporating Student Interaction	
Please rate the professor's ability to provide students with an opportunity to participate.	4.00
Please rate the professor's ability to ask questions that deepened students' understanding of the course material.	1.00
Please rate the professor's ability to allow students to provide meaningful contributions.	2.00
Overall Rating	
Please rate the professor's overall performance.	3.00

CHAPTER III: RESULTS

Descriptive statistics and frequency counts were calculated for all qualitative questions related to the participants' demographic information. To assess the posed hypotheses and research questions, a series of descriptive statistics, Fisher R to Z correlation transformations and z test, and between-subjects MANOVA were computed to determine the accuracy differences between BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples. The results were analyzed for the 13 performance evaluations, otherwise referred to as composite ratings, and the overall ratings for each course. The scale alphas were calculated for BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples. Additionally, the absolute differences, as well as over-rating and under-rating, were computed using the expert rater scores.

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

Hypothesis one was tested by comparing the scale alphas of the composite ratings for each course between BARS with five anchors/examples and BARS with three anchors/examples. When calculating the scale alphas, the researchers observed that the reliability estimates would be enhanced across all six courses if the performance evaluation rating, "Please rate the professor's ability to draw upon scholarly works," was removed from the set of 13 performance evaluation ratings. Therefore, the item was removed prior to reporting the scale alphas found in Table 7.

The results were significant for the Introduction to Psychology composite ratings, $z = -2.008$, $p < .05$. The results were nonsignificant for the Introduction to Philosophy

composite ratings, $z = 0.174$, $p > .05$. The results were nonsignificant for the Introduction to World History composite ratings, $z = 0.460$, $p > .05$. The results were nonsignificant for the Developmental Psychology composite ratings, $z = 0.566$, $p > .05$. The results were significant for the Computer Science composite ratings, $z = 2.231$, $p < .05$. The results were nonsignificant for the Physics composite ratings, $z = 0.223$, $p > .05$. For overall ratings, the results were nonsignificant, $z = 1.326$, $p > .05$. The results did not support hypothesis 1a for all lectures/professors.

Hypothesis one was tested by comparing the absolute differences between composite ratings and true score estimates among BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples. A between-subjects MANOVA was conducted to test whether BARS with five behavioral anchors/examples resulted in lower absolute differences between composite ratings and true score estimates compared to BARS with three behavioral anchors/examples. The results were nonsignificant for composite ratings, $F(6, 70) = 1.039$, $p = .408$. The results were also nonsignificant for the overall ratings, $F(6, 70) = 2.139$, $p = .059$. The results did not support hypothesis 1b.

Lastly, hypothesis one was tested using a between-subjects MANOVA to compare the amount of over-rating and under-rating, as compared to true score estimates, between BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples. The results were nonsignificant for composite ratings, $F(6, 70) = 1.906$, $p = .092$. The results were also nonsignificant for the overall ratings, $F(6, 70) = 0.211$, $p = .972$. The results did not support hypothesis 1c.

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

Scale alphas were calculated for the 13 performance rating items and compared between participants who scored relatively high on conscientiousness (scores above 3.34) and those that scored relatively low on conscientiousness (scores below 3.33). When calculating the scale alphas, the researchers observed that the reliability estimates would be enhanced across all courses if the performance evaluation rating, "Please rate the professor's ability to draw upon scholarly works," was removed from the set of 13 performance evaluation ratings. Therefore, the item was removed prior to reporting the scale alphas found in Table 8.

The scale alphas for the two groups were analyzed using a Fisher's r to z transformation. Results were nonsignificant for the Introduction to Psychology composite ratings, $z = -0.048$, $p > .05$. The results were nonsignificant for the Introduction to Philosophy composite ratings, $z = 0.502$, $p > .05$. The results were significant for the Introduction to World History composite ratings, $z = 2.955$, $p < .05$. Further investigation of this significant finding revealed that all participants who scored high on conscientiousness consistently rated the professor a one on the evaluation rating, "Please rate the professor's ability to ask questions that deepened students' understanding of the course material." The results were nonsignificant for the Developmental Psychology composite ratings, $z = 0.515$, $p > .05$. The results were significant for the Computer Science composite ratings, $z = 0.631$, $p < .05$. Further investigation of this significant finding showed that all participants who scored high on conscientiousness consistently rated the professor a five on the evaluation rating, "Please rate the professor's ability to

communicate information with or without the use of filler words." The results were nonsignificant for the Physics composite ratings, $z = -0.811$, $p > .05$. For overall ratings, the results were nonsignificant, $z = -0.450$, $p > .05$. The results did not support the first research question for all lectures/professors.

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

A repeated measures within-subjects MANOVA was conducted to see whether absolute differences vary for composite ratings based on rater conscientiousness between BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples. The results showed no significant differences for composite ratings $F(6, 69) = 1.005$, $p = .429$, as well as for the overall ratings, $F(6, 69) = 2.094$ $p = .065$.

Another repeated measures within subject MANOVA was conducted to see whether over-under ratings differ for composite ratings based on rater conscientiousness between BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples. The results revealed no significant differences in the accuracy of composite ratings, $F(6, 69) = 1.842$, $p = .104$, as well as for the overall ratings $F(6, 69) = 0.179$, $p = .982$. The results did not support the second research question.

Table 7
Fisher R to Z Correlation Transformations for BARS
Outcome Variable: Rater Accuracy

Scale	r of 3 examples	r of 5 examples	n of 3 examples	n of 5 examples	Fisher's z
Introduction to Psychology	0.679	0.863	36	41	-2.008
Introduction to Philosophy	0.902	0.894	36	41	0.174
Introduction to World History	0.945	0.932	36	41	0.460
Developmental Psychology	0.907	0.880	36	41	0.566
Computer Science	0.942	0.841	36	41	2.231
Physics	0.865	0.851	36	41	0.223
Overall Rating	0.698	0.499	36	41	1.326

Note: Bolded z values are significant at $p < .05$

Table 8
Fisher R to Z Correlation Transformations for Conscientiousness
Outcome Variable: Rater Accuracy

Scale	r of low C	r of high C	n of 3 low C	n of high C	Fisher's z
Introduction to Psychology	0.773	0.779	29	20	-0.048
Introduction to Philosophy	0.885	0.846	29	20	0.502
Introduction to World History	0.962	0.782	29	20	2.955
Developmental Psychology	0.913	0.882	29	20	0.515
Computer Science	0.909	0.868	29	20	0.631
Physics	0.866	0.917	29	20	-0.811
Overall Rating	0.704	0.768	29	20	-0.450

Note: Bolded z values are significant at $p < .05$; C = conscientiousness.

CHAPTER IV: DISCUSSION

Previous research has focused on understanding what type of measurement tool produces the best accuracy estimates. The present study sought to determine whether using BARS with five behavioral anchors/examples or three behavioral anchors/examples impacted the accuracy of ratings. The results of this study demonstrated the importance of defining behavior and anchoring the exemplars on a scale from least to most effective in order to yield ratings that are more accurate.

Hypothesis one stated that BARS with five anchors/examples would be more accurate in scoring performance evaluations than BARS with three anchors/examples. The hypothesis was parsed out into three analyses to determine whether BARS with five anchors/examples would result in higher scale alphas, lower absolute differences, and less over-rating and under-ratings, relative to BARS with three anchors/examples. The results showed that the accuracy of performance evaluation ratings did not significantly differ between BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples in regards to scale alphas, absolute differences, and over-under rating. This nonsignificant finding suggests that the accuracy of performance evaluation ratings will not be impacted by the number of behavioral anchors/examples when comparing BARS with five behavioral anchors/examples to BARS with three behavioral anchors/examples. While Iseral (2020) found support for BARS with five behavioral anchors/examples when assessing over-under ratings, the present study did not find significant differences. This could be due to the design of both studies. Iseral (2020) asked participants to evaluate a single candidate, whereas the current study requested participations provide performance evaluation ratings for six professors.

In addition to testing the hypotheses, the researchers also analyzed two research questions. The first research question investigated whether there were differences in scale alphas of the ratings based upon the degree of rater conscientiousness. The results of this study revealed that there are no significant differences in scale alphas based upon the degree of rater conscientiousness between BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples. However, two of the six courses demonstrated significant differences between rater conscientiousness and the accuracy of ratings. In other words, participants who demonstrated a high degree of conscientiousness provided consistent ratings for a single performance evaluation rating. Future research is needed to understand if the degree of rater conscientiousness yields different scale alphas based on the number of behavioral anchors/examples.

The second research question investigated whether accuracy estimates, both absolute value and over-under ratings, vary for ratings based upon rater conscientiousness. The results of this study demonstrated that there are no significant differences between the degree of rater conscientiousness and the accuracy of ratings based BARS with five behavioral anchors/examples and BARS with three behavioral anchors/examples.

Theoretical Implications

The present study provides a deeper understanding of how the number of behavioral anchors/examples presented on BARS impacts the accuracy of ratings. The results showed that when comparing scale alphas of five behavioral anchors/examples and three behavioral anchors/examples, there were no significant differences. Future research should continue to investigate the interaction between the number of behavioral

anchors/examples and the accuracy of performance evaluation ratings. The present study asked participants to evaluate the performance of six professors using BARS, whereas Iseral (2020) requested participants to assess a single candidate. Future research should consider evaluating the performance of a single ratee to observe if differences in accuracy estimates arise. In addition, future research should consider incorporating rater training as it has been shown to effectively reduce halo and leniency (Borman, 1979).

Future research could also investigate whether combining the performance evaluation ratings for one dimension onto a single scale impacts the accuracy estimates of BARS with five behavioral anchors/examples compared to BARS with three behavioral anchors/examples. For instance, BARS in the present study presented an independent scale for each performance evaluation rating, resulting in 13 performance evaluation rating scales. By combining all performance evaluation ratings onto a single scale per dimension, raters could perceive ratee performance differently, thus impacting rater accuracy estimates. Prior research claims that ratees may be more inclined to accept the results of a performance evaluation if they participate in the development of the BARS (Schwab et al., 1975). Future research could also assess ratee acceptance of results or feedback when using BARS with five behavioral anchors/examples versus three behavioral anchors/examples. The present study did not observe consistent differences in rater accuracy as a function of rater conscientiousness. Future research should continue to evaluate how the degree of rater conscientiousness interacts with the accuracy of performance evaluation ratings based on the number of behavioral anchors/examples.

Practical Implications

While the results of the study were not what the researchers had hypothesized, relevant practical considerations still exist based on the results. It is beneficial for practitioners to understand that accuracy estimates do not differ for performance evaluations when using BARS with five behavioral anchors/examples versus BARS three behavioral anchors/examples. By only constructing three behavioral anchors/examples, organizations could conserve their resources when developing a performance management system that involves BARS. However, it is imperative that ratees remain involved in the process of creating BARS to maintain acceptance and understanding of the human resource tool (Schwab et al., 1975). If practitioners are curious about the interaction of personality and the accuracy of performance evaluation ratings, the degree of rater conscientiousness does not impact accuracy estimates, according to the present study. However, further research is necessary to understand whether conscientiousness does in fact interact with the number of behavioral anchors/examples and rater accuracy. The results demonstrated that practitioners could use BARS with three behavioral anchors/examples for performance evaluations and observe accurate ratings.

Limitations

While two manipulation check questions were included near the end of the study, participants did not meet the minimum criteria in order to retain their data. The first manipulation check question asked participants to indicate how many behavioral anchors/examples they were presented with and what side of the screen had behavioral anchors/examples of good performance versus bad performance. Participants often responded with the appropriate number of behavioral anchors/examples, but they often

failed to respond correctly to the orientation of the scale from bad performance to good performance. The second manipulation check question yielded a better response rate compared to the first check, however not enough participants passed to conduct analyses. Therefore, the researchers relied on participant responses to the attention check questions, specifically the course subject and professor gender. As a result, the final sample size of participants may not have been sufficient in order to analyze the results and generate enough power to determine if the posed hypotheses and research questions would yield significant findings. With a larger sample size, future research could identify significant findings for over-under rating, similar to Iseral (2020), as well as significant findings for absolute differences.

Conclusion

Performance evaluations are designed to assess how well an individual is performing the job using a rating format that yields accurate ratings. Past research indicates that BARS is one of the standardized measurement tools which yields accuracy estimates (Burnaska & Hollmann, 1974; Campbell et al., 1973; Dickinson & Zellinger, 1980; Schwab et al., 1975). Thus, the present study explored the differences in accuracy ratings based on the number of behavioral anchors/examples. While BARS commonly includes three behavioral anchors/examples, the researchers sought to investigate if adding two additional anchors/examples would increase accuracy. The researchers did not find differences in the accuracy of performance evaluation ratings when comparing BARS with five behavioral anchors/examples to BARS with three behavioral anchors/examples. The study also showed no differences in the accuracy of ratings based on the degree of rater conscientiousness when using BARS with five behavioral

anchors/examples versus BARS with three behavioral anchors/examples. Researchers should continue to study how to best construct performance evaluation measurement tools to ensure equitable human resource decisions are made for all employees.

REFERENCES

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. Journal of Applied Psychology, 2.

Bernardin, H. J., & Smith, P. C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored ratings scales (BARS). Journal of Applied Psychology, 66(4), 458–463.

Borman, W. C. (1974). The ratings of individuals in organizations: An alternate approach. Organizational Behavior and Human Performance, 12, 105- 124.

Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. Organizational Behavior and Human Performance, 20(2), 238-252.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63(2).

Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64(4), 410–421.

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.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO personality inventory and the NEO five-factor inventory. Odessa, FL: Psychological Assessment Resources.

DeNisi, A. S., & Murphy, K. R. (2017). Performance appraisal and performance management: 100 years of progress? Journal of Applied Psychology, 102(3), 421–433.

DeNisi, A. S., & Pritchard, R. D. (2006). Performance appraisal, performance management and improving individual performance: a motivational framework. Management and Organization Review, 2, 253.

Dickinson, T. L., & Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. Journal of Applied Psychology, 65(2), 147–154.

Flanagan, J.C. (1954). The critical incident technique. Psychological Bulletin, 51(4), 327-358.

Harrell, A., & Wright, A. (1990). Empirical evidence on the validity and reliability of behaviorally anchored rating scales for auditors. Auditing: A Journal of Practice & Theory, 9(3), 134–149.

Hauenstein, N. M., Brown, R. D., & Sinclair, A. L. (2010). BARS and those mysterious, missing middle anchors. Journal of Business and Psychology, 25(4), 663-672.

Kingstrom, P. O., & Bass, A. R. (1981). A critical analysis of studies comparing behaviorally anchored rating scales (BARS) and other rating formats. Personnel Psychology, 34(2), 263–289.

Latham, G. P., Almost, J., Mann, S., & Moore, C. (2005). New developments in performance management. Organizational Dynamics, 1, 77.

Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: a Monte Carlo approach. Journal of Applied Psychology, 60(1), 10–13.

Lunenburg, F. (2012). Performance appraisal: Methods and rating errors. International Journal of Scholarly Academic Intellectual Diversity, 14(1), 1-9.

Murphy, K. R., & Cleveland, J. (1995). Understanding performance appraisal: Social, organizational, and goal-based perspectives. Sage Publications.

Murphy, K. R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W.K. (1982). Relationship between observational accuracy and accuracy in evaluating performance. Journal of Applied Psychology, 67, 320- 325.

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.

Roch, S. G., Ayman, R., Newhouse, N., & Harris, M. (2005). Effect of identifiability, rating audience, and conscientiousness on rating level. International Journal of Selection and Assessment, 13(1), 53–62.

Schwab, D. P., Heneman, I. H. G., & DeCotiis, T. A. (1975). Behaviorally anchored rating scales: a review of the literature. Personnel Psychology, 28(4), 549–562.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149–155.

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. Journal of Applied Psychology, 73(3), 497–506.

Tziner, A., Murphy, K. R., & Cleveland, J. N. (2002). Does conscientiousness moderate the relationship between attitudes and beliefs regarding performance appraisal and

rating behavior? International Journal of Selection and Assessment, 10(3), 218-224.

Wilson, M. A., Bennett, W., Gibson, S.G., & Alliger, G. M. (2017). The handbook of work analysis: The methods, systems, applications and science of work measurement in organizations. Routledge.

APPENDICES

APPENDIX A: PERFORMANCE EVALUATION DIMENSIONS AND DEFINITIONS

Presentation Ability
Please rate the professor's ability to show excitement about the course.
Please rate the professor's ability to use presentation aids.
Please rate the professor’s ability to communicate information about the course.
Please rate the professor’s ability to communicate information with or without the use filler words.
Organization and Clarity
Please rate the professor's ability to present the course material in an organized, logical framework.
Please rate the professor’s ability to teach at a pace that supports learning.
Please rate the professor’s ability to teach at a level that supports learning.
Please rate the professor's ability to make an effort to gauge students' understanding of the course material.
Scholarly and Intellectual
Please rate the professor's ability to focus on relevant course material.
Please rate the professor's ability to draw upon scholarly works.
Incorporating Student Interaction
Please rate the professor's ability to provide students with an opportunity to participate.
Please rate the professor's ability to ask questions that deepened students' understanding of the course material.
Please rate the professor's ability to allow students to provide meaningful contributions.
Overall Rating
Please rate the professor's overall performance.

APPENDIX B: BARS WITH FIVE BEHAVIORAL ANCHORS/EXAMPLES

Presentation Ability

Please rate the professor's ability to show excitement about the course.

	Did not show excitement about the course	Showed little excitement about the course	Showed some excitement about the course but not enough to hold the interest of the students	Showed excitement about the course to hold the interest of the students some of the time	Showed a lot of excitement about the course to hold the interest of the students most of the time
Presentation Ability	○	○	○	○	○

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

	Did not use presentation aids	Used one type of presentation aid and it did not enhance student learning	Used one type of presentation aid and it somewhat enhanced student learning	Used one type of presentation aid and it enhanced student learning	Used more than one type of presentation aid and it really enhanced student learning
Presentation Ability	○	○	○	○	○

Please rate the professor's ability to communicate information about the course.

	Did not communicate information about the course	Communicated information about the course poorly	Communicated information about the course somewhat well	Communicated information about the course well	Communicated information about the course exceptionally well
Presentation Ability	○	○	○	○	○

Please rate the professor's ability to communicate information with or without the use filler words.

	Constantly used filler words such as "um" or "like"	Typically used filler words such as "um" or "like"	Often used filler words such as "um" or "like"	Occasionally used filler words such as "um" or "like"	Rarely used filler words such as "um" or "like"
Presentation Ability	○	○	○	○	○

Organization and Clarity

Please rate the professor's ability to teach at a level that supports learning.

	Did not teach at a level that supported learning and never allowed students to reflect	Taught at a poor level that did not support learning and rarely allowed students to reflect	Taught at a decent level that supported learning and occasionally allowed students to reflect	Taught at a good level that supported learning and often allowed students to reflect	Taught at an exceptional level that supported learning and typically allowed students to reflect
Organization and Clarity	○	○	○	○	○

Please rate the professor's ability to make an effort to gauge students' understanding of the course material.

	Did not make an effort to gauge students' understanding of course material	Made a little effort to gauge students' understanding of course material	Made some effort to gauge students' understanding of course material by asking a question	Made some effort to gauge students' understanding of course material by asking a question and scanning the room	Made many efforts to gauge students' understanding of course material by asking numerous questions and scanning the room
Organization and Clarity	○	○	○	○	○

Please rate the professor's ability to present the course material in an organized, logical framework.

	Did not present the course material in an organized or logical framework	Presented the course material in an organized or logical framework	Presented the course material in a somewhat organized, logical framework but did not provide background information regarding the topic	Presented the course material in an organized, logical framework by providing background information regarding the topic or clarifying what will be discussed in class	Presented the course material in a clearly organized, logical framework by providing background information regarding the topic and clarifying what will be discussed in class
Organization and Clarity	○	○	○	○	○

Please rate the professor's ability to teach at a pace that supports learning.

	Did not teach at a pace that supported learning and never allowed students to ask questions	Taught at a poor pace that did not support learning and rarely allowed students to ask questions	Taught at a decent pace that supported learning and occasionally allowed students to ask questions	Taught at a good pace that supported learning and often allowed students to ask questions	Taught at an exceptional pace that supported learning and typically allowed students to ask questions
Organization and Clarity	○	○	○	○	○

Intellectual and Scholarly

Please rate the professor's ability to focus on relevant course material.

	Did not focus on relevant course material	Focused a little bit of the time relevant course material	Focused some of the time on relevant course material but provided irrelevant examples	Focused most of the time on relevant course material and provided a relevant example	Focused all of the time on relevant course material and provided relevant examples
Intellectual and Scholarly	○	○	○	○	○

Please rate the professor's ability to draw upon scholarly works.

	Did not draw upon scholarly works regarding the topic	Drew upon a scholarly work regarding the topic	Drew upon a scholarly work regarding the topic but did not provide an example	Drew upon a scholarly work regarding the topic and provided an example	Drew upon more than one scholarly work regarding the topic and provided some examples
Intellectual and Scholarly	○	○	○	○	○

Incorporating Student Interaction

Please rate the professor's ability to provide students with an opportunity to participate.

	Did not provide students with an opportunity to participate	Provided a student with an opportunity to participate	Provided some students with an opportunity to participate but did not respond to students	Provided most students with an opportunity to participate and responded to some students	Provided most students with an opportunity to participate and responded to each student
Incorporating Student Interaction	○	○	○	○	○

Please rate the professor's ability to ask questions that deepened students' understanding of the course material.

	Did not ask questions to deepen students' understanding of the course material	Asked a question to deepen students' understanding of the course material	Asked a few questions to deepen students' understanding of the course material but does not help clarify their thinking	Asked some questions to deepen students' understanding of the course material and helps clarify their thinking	Asked many questions to deepen students' understanding of the course material and helps clarify their thinking
Incorporating Student Interaction	○	○	○	○	○

Please rate the professor's ability to allow students to provide meaningful contributions.

	Did not allow students to provide contributions to the lecture	Allowed students to provide contributions to the lecture	Allowed students to provide contributions to the lecture by engaging them in the thought process	Allowed students to provide meaningful contributions to the lecture by engaging them in the thought process	Allowed and encouraged students to provide meaningful contributions to the lecture by engaging them in the thought process
Incorporating Student Interaction	○	○	○	○	○

APPENDIX C: BARS WITH THREE BEHAVIORAL ANCHORS/EXAMPLES

Presentation Ability

Please rate the professor's ability to show excitement about the course.

	Did not show excitement about the course		Showed some excitement about the course but not enough to hold the interest of the students		Showed a lot of excitement about the course to hold the interest of the students most of the time
Presentation Ability	○	○	○	○	○

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

	Did not use presentation aids		Used one type of presentation aid and it somewhat enhanced student learning		Used more than one type of presentation aid and it really enhanced student learning
Presentation Ability	○	○	○	○	○

Please rate the professor's ability to communicate information about the course.

	Did not communicate information about the course		Communicated information about the course somewhat well		Communicated information about the course exceptionally well
Presentation Ability	○	○	○	○	○

Please rate the professor's ability to communicate information with or without the use filler words.

	Constantly used filler words such as "um" or "like"		Often used filler words such as "um" or "like"		Rarely used filler words such as "um" or "like"
Presentation Ability	○	○	○	○	○

Organization and Clarity

Please rate the professor's ability to present the course material in an organized, logical framework.

	Did not present the course material in an organized or logical framework		Presented the course material in a somewhat organized, logical framework but did not provide background information regarding the topic		Presented the course material in a clearly organized, logical framework by providing background information regarding the topic and clarifying what will be discussed in class
Organization and Clarity	○	○	○	○	○

Please rate the professor's ability to teach at a pace that supports learning.

	Did not teach at a pace that supported learning and never allowed students to ask questions		Taught at a decent pace that supported learning and occasionally allowed students to ask questions		Taught at an exceptional pace that supported learning and typically allowed students to ask questions
Organization and Clarity	○	○	○	○	○

Please rate the professor's ability to teach at a level that supports learning.

	Did not teach at a level that supported learning and never allowed students to reflect		Taught at a decent level that supported learning and occasionally allowed students to reflect		Taught at an exceptional level that supported learning and typically allowed students to reflect
Organization and Clarity	○	○	○	○	○

Please rate the professor's ability to make an effort to gauge students' understanding of the course material.

	Did not make an effort to gauge students' understanding of course material		Made some effort to gauge students' understanding of course material by asking a question		Made many efforts to gauge students' understanding of course material by asking numerous questions and scanning the room
Organization and Clarity	○	○	○	○	○

Intellectual and Scholarly

Please rate the professor's ability to focus on relevant course material.

	Did not focus on relevant course material		Focused some of the time on relevant course material but provided irrelevant examples		Focused all of the time on relevant course material and provided relevant examples
Intellectual and Scholarly	○	○	○	○	○

Please rate the professor's ability to draw upon scholarly works.

	Did not draw upon scholarly works regarding the topic		Drew upon a scholarly work regarding the topic but did not provide an example		Drew upon more than one scholarly work regarding the topic and provided some examples
Intellectual and Scholarly	○	○	○	○	○

Incorporating Student Interaction

Please rate the professor's ability to provide students with an opportunity to participate.

	Did not provide students with an opportunity to participate		Provided some students with an opportunity to participate but did not respond to students		Provided most students with an opportunity to participate and responded to each student
Incorporating Student Interaction	○	○	○	○	○

Please rate the professor's ability to ask questions that deepened students' understanding of the course material.

	Did not ask questions to deepen students' understanding of the course material		Asked a few questions to deepen students' understanding of the course material but does not help clarify their thinking		Asked many questions to deepen students' understanding of the course material and helps clarify their thinking
Incorporating Student Interaction	○	○	○	○	○

Please rate the professor's ability to allow students to provide meaningful contributions.

	Did not allow students to provide contributions to the lecture		Allowed students to provide contributions to the lecture by engaging them in the thought process		Allowed and encouraged students to provide meaningful contributions to the lecture by engaging them in the thought process
Incorporating Student Interaction	○	○	○	○	○

APPENDIX D: ATTENTION CHECK QUESTIONS

1. What was the course subject?
Psychology
Philosophy
World History
Developmental Psychology
Computer Science
Physics

2. Which topic was discussed during the lecture?
Psychology
Introduction
Personality
Abnormal behavior
Philosophy
Truth and knowledge
Good and bad teaching
Mind and body
World History
Introduction
Holocaust
Civil Rights Movement
Developmental Psychology
Perception
Lifespan
Gender Roles
Computer Science
Programming
Cyber Security
Algorithms
Physics
Introduction
Gravity
Electricity

3. What was the professor's gender?
Male
Woman
4. Did the professor walk around during the lecture?
Yes
No
5. Did the professor sit on the desk during the lecture?
Yes
No
6. Was the professor wearing glasses?
Yes
No

APPENDIX E: MANIPULATION CHECK QUESTIONS

1. The rating scale you used was on a scale of:

Five (5) options where good performance was on the left side of the screen and bad performance was on the right side of the screen

Five (5) options where good performance was on the right side of the screen and bad performance was on the left side of the screen

Three (3) options where good performance was on the left side of the screen and bad performance was on the right side of the screen

Three (3) options where good performance was on the right side of the screen and bad performance was on the left side of the screen

2. The rating scales you used to conduct professor evaluations had:

Two (2) behavioral examples for each rating provided

Three (3) behavioral examples for each rating provided

Four (4) behavioral examples for each rating provided

Five (5) behavioral examples for each rating provided

APPENDIX F: HEXACO

60 item index with a 5-point Likert scale and the following scale points:

1 = Strongly Disagree
2 = Disagree
3 = Neither Agree nor Disagree
4 = Agrees
5 = Strongly Agree

1. I would be quite bored by a visit to an art gallery.
2. I plan ahead and organize things, to avoid scrambling at the last minute.
3. I rarely hold a grudge, even against people who have badly wronged me.
4. I feel reasonably satisfied with myself overall.
5. I would feel afraid if I had to travel in bad weather conditions.
6. I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.
7. I'm interested in learning about the history and politics of other countries.
8. I often push myself very hard when trying to achieve a goal.
9. People sometimes tell me that I am too critical of others.
10. I rarely express my opinions in group meetings.
11. I sometimes can't help worrying about little things.
12. If I knew that I could never get caught, I would be willing to steal a million dollars.
13. I would enjoy creating a work of art, such as a novel, a song, or a painting.
14. When working on something, I don't pay much attention to small details.
15. People sometimes tell me that I'm too stubborn.
16. I prefer jobs that involve active social interaction to those that involve working alone.
17. When I suffer from a painful experience, I need someone to make me feel comfortable.
18. Having a lot of money is not especially important to me.
19. I think that paying attention to radical ideas is a waste of time.
20. I make decisions based on the feeling of the moment rather than on careful thought.
21. People think of me as someone who has a quick temper.
22. On most days, I feel cheerful and optimistic.
23. I feel like crying when I see other people crying.
24. I think that I am entitled to more respect than the average person is.
25. If I had the opportunity, I would like to attend a classical music concert.
26. When working, I sometimes have difficulties due to being disorganized.

27. My attitude toward people who have treated me badly is “forgive and forget”.
28. I feel that I am an unpopular person.
29. When it comes to physical danger, I am very fearful.
30. If I want something from someone, I will laugh at that person's worst jokes.
31. I’ve never really enjoyed looking through an encyclopedia.
32. I do only the minimum amount of work needed to get by.
33. I tend to be lenient in judging other people.
34. In social situations, I’m usually the one who makes the first move.
35. I worry a lot less than most people do.
36. I would never accept a bribe, even if it were very large.
37. People have often told me that I have a good imagination.
38. I always try to be accurate in my work, even at the expense of time.
39. I am usually quite flexible in my opinions when people disagree with me.
40. The first thing that I always do in a new place is to make friends.
41. I can handle difficult situations without needing emotional support from anyone else.
42. I would get a lot of pleasure from owning expensive luxury goods.
43. I like people who have unconventional views.
44. I make a lot of mistakes because I don’t think before I act.
45. Most people tend to get angry more quickly than I do.
46. Most people are more upbeat and dynamic than I generally am.
47. I feel strong emotions when someone close to me is going away for a long time.
48. I want people to know that I am an important person of high status.
49. I don’t think of myself as the artistic or creative type.
50. People often call me a perfectionist.
51. Even when people make a lot of mistakes, I rarely say anything negative.
52. I sometimes feel that I am a worthless person.
53. Even in an emergency I wouldn’t feel like panicking.
54. I wouldn’t pretend to like someone just to get that person to do favors for me.
55. I find it boring to discuss philosophy.
56. I prefer to do whatever comes to mind, rather than stick to a plan.
57. When people tell me that I’m wrong, my first reaction is to argue with them.
58. When I’m in a group of people, I’m often the one who speaks on behalf of the group.
59. I remain unemotional even in situations where most people get very sentimental.
60. I’d be tempted to use counterfeit money, if I were sure I could get away with it.

APPENDIX G: IRB APPROVAL FORM

Human Participant Research Proposal
IRBF004: EXEMPTION REQUEST FORM

"Exempt" Definition:
It is important that seekers of IRB exemption understand that "exempt" does not reflect its literal meaning but those protocols that qualify for "exempt status" are often reviewed by the MTSU Office of Compliance and do not require an annual continuing review. However, the procedure and documents requirement for exempt protocols are mostly same in comparison to those protocols that require more IRB oversight.

What does this form contain?
This new exemption request form contains several newly added features to help researchers to clearly outline their proposal to collect data from living individuals. Although more information is requested from the applicants, the review process is expected to focus on the research and human intervention than on minor issues. This form also contains space for reviewer comments thereby allowing the review process to resemble an informative discussion. The applicant must provide the necessary details for questions in Sections 1-11 (Refer to the following list of contents). The Sections 12 & 13 are for Office Use only.

1. Project Information
2. Investigator Information
3. Exemption Determination
4. Exemption for Research with minors
5. Selection of Research Category
6. Research Methods & Instruments
7. Participant Selection & Recruitment
8. Informed Consent
9. CITI Training
10. Mandatory Documents & Attachments
11. Investigators' Declaration and Assurance
12. *Review (Office Use)*
13. *IRB Action (Office Use)*

Mandatory requirements

- Completed informed consent form - Click
- All of the investigators must complete all required research-specific CITI training modules
- Provide a detailed strategy for avoiding COVID-19 infection if the participants will have direct interaction
- In addition, other documents may be required

Instructions for document submission.

- This application and support documents must be submitted by the faculty member who signs Section 11.2.
- Send all documents as separate files but in a single email to irb_submissions@mtsu.edu
- Submit all IRB forms in their original MS Word format – DO NOT CONVERT TO PDF

Review & Timeline

- Once the OC confirms that the application is complete, a complete review will be completed within 2 weeks
- This form will be sent back to the investigators with reviewers' comments and other instructions
- The review process is iterative and it depends on how swiftly the investigators are able to address all reviewers' concerns.
- Once a final approval has been issued, a "locked" version of this form will be sent to the investigators to be used as a guideline for their study.

This form also contains space for reviewer comments. Therefore, do not convert this to PDF but instead send the completed form to irb_submissions@mtsu.edu in its original MS Word format.

1. PROJECT INFORMATION

1.1 **Choose your review type:** ☒ EXEMPT
Review

1.2 Enter Project Title

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

1.3 Primary Investigator or Principal Investigator (PI) Information:

☐Faculty[4] ☐Staff[4] ☒Graduate[5,6] ☐Undergraduate[5,6] ☐Other[5,6]

Name	Kelsey Byerly
Email	krb7r@mtmail.mtsu.edu *Telephone:* 407-341-8226
Alternate Email	kelsbyerly@gmail.com **if PI is a student*
Department/Unit	Psychology *College* College of Behavioral and Health Sciences
Office Location	*Room #* *Building* *Box #*
Contact Address	1955 Old Castle Dr, Apt 123, Murfreesboro, TN 37127
CITI Program ID	7015212

Refer to https://www.mtsu.edu/irb/FAQ/ResponsibilitiesOfPI.php for PI responsibilities.

1.1 Faculty Advisor (FA) if the PI is a student:

Name	Dr. Makr Frame ☒Faculty ☐Staff ☐Other
Email	mark.frame@mtsu.edu *Telephone:* 615-898-2565
Department/Unit	Psychology *College* College of Behavioral and Health Sciences
Office Location	*Room* #251 *Building* Academic Classroom Building *Box* #87
CITI Program ID	1963131

Refer https://www.mtsu.edu/irb/FAQ/Faculty.php

- Must be completed by an MTSU faculty or a FTE if the PI is a student.
- The FA must submit the application packet by email to *irb_submissions@mtsu.edu* indicating that s/he has knowledge of this proposal.

1.4 Investigating Team *(List ALL Investigators - contact the Compliance Office for more than 6 Co-I's)*

Name/Email/Status	Department/Affiliation	CITI Training
Name: Dr. Michael Hein Email: michael.hein@mtsu.edu ☒Faculty/Staff ☐Student[5] ☐Non-MTSU ☐Other	Psychology	☒ Completed
Name: Dr. Alex Jackson Email: alexander.jackson@mtsu.edu ☒Faculty/Staff ☐Student[5] ☐Non-MTSU ☐Other	Psychology	☒ Completed

Foot Notes:

[4] Faculty PI must complete and sign Sections 11.1 and 11.2

[5] The Student PI must complete Section 11.1 and an MTSU Faculty Advisor/mentor must sign Section 11.2. In addition, the application documents MUST be emailed to irb_submissions@mtsu.edu by the MTSU Faculty who completes Section 10.2 with a statement of approval in the body of the email.

[6] The Students, regardless of their affiliation, MUST complete "Students in Research" module from CITI Program

[7] The faculty advisor or sponsor MUST be an MTSU faculty member.

1.5 Submission Status of this Study:

☒ New Submission[1] ☐ Revision[2] ☐ Previous Protocol ID(s) given to this study[3]

[1] Check this box if this is the first time you are submitting this study for IRB review

[2] Check this box if you have already submitted this application to the IRB but you have been asked to make revisions to your application or other documents by the IRB or by the Compliance Staff

[3] Check this box and provide the IRB ID if you are trying to extend a previously approved IRB protocol

1.6 Research Classification *(select ALL that apply):*

☒ Social/Behavioral/Educational Research ☐ Biomedical Research

☐ Clinical Research ☐ Quality Assurance/Evaluation

1.7 Research Category *(select ALL that apply):*

☐ Faculty/Staff research ☐ FRCAC ☐ URECA ☐ Class Project

☒ Thesis Dissertation ☐ Not for Publication ☐ Publication/Presentation

☐ Other

1.8 Miscellaneous Questions:

Project Questions	**Response**	**Remark(s)**
Expected start date	03/12/2021	
Anticipated completion date The protocol will be closed on this date	12/31/2023	
Source of funding (Funding agency, number/ID, and expiration date)	NA	

Review Tracking

Protocol ID	**21-1131 2q**	**IRB Comments**
Date Received	**03/04/2021**	
Prescreen Date	**03/08/2021**	
Revision Date (if applicable)	**N/R**	
Review Date	**03/11/2021**	
Revision Date (if applicable)	**N/R**	Administrative revision of the recruitment scripts, informed consent and etc.
Approval Date	**03/11/2021**	
Expiration Date	**06/30/2024**	This expiration date may be extended by making a written request.

2. EXEMPT DETERMINATION QUESTIONAIRE

2.1	**Vulnerable Subjects -** Are the subjects from a vulnerable group, such as, prisoners, seriously ill, cognitively impaired, protected minorities and/or *etc.*?	☐ **Yes** ☒ No
2.2	**Risk to the Subjects -** Does the research involve the collection of behavioral data which, if known outside the research, could reasonably place the subjects at risk for criminal or civil liabilities or be damaging to the individual's financial standing, employability or reputation?	☐ **Yes** ☒ No
2.3	**Sensitive Topics -** Will you be collecting information regarding sensitive topics or personal aspects of a subject's behavior, such as, drug or alcohol use, illegal conduct, sexual behavior, mental health an/or etc.?	☐ **Yes** ☒ No
2.4	**Video/audio -** Will you be audio/video recording participant's response?	☐ **Yes** ☒ No
2.5	**Discomfort(s) to the Subjects -** Will this study expose the subjects to discomfort or stress beyond the levels encountered in daily life?	☐ **Yes** ☒ No
2.6	**Research with Minors -** Does your research involve *collection of data from minors* or *use of data collected previously from minors*? Complete Section 4 if Yes	☐ **Yes** ☒ No

Other than question 3.6, if you answered "YES" to any of the above questions, then t research is DISQUALIFIED from obtaining an exempt designation

3. RESEARCH WITH MINORS

NONE

4. RESEARCH CATEGORIES

The Federal Code [45 CFR 46 (46.101)] identifies the activities that fall within the following six categories as exempt. You MUST select the appropriate exemption category that apply to this study.

	Exemption Category - research activities that are exempt from continuing review	
1	Research conducted in established or commonly accepted **educational settings**, involving normal educational practices, such as, (i) research on regular and special education **instructional strategies**, or (ii) research on effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods	☐
2	Research involving the use of **educational tests** (cognitive diagnostic, aptitude, achievement), **survey procedures, interviews or observation of public behavior**, UNLESS (i) information obtained is recorded in such a manner that human subjects can be identified directly or through identifiers linked to the subjects; AND (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk or criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation	☒
3	Research involving the use of **educational tests** (cognitive diagnostic, aptitude, achievement), **survey procedures, interviews or observation of public behavior** that is not exempt in 5.2 of this section if: (i) the human subjects are elected or **appointed public officials or candidates for public office**; OR (ii) Federal statute(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.	☐
4	Research involving the collection or study of **existing data, documents, records** (pathological specimens or diagnostic specimens) if publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects or the data were collected through a different protocol approved by an ethics committee such as the IRB	☐
5	Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate or otherwise examine: (i) **Public benefit or service programs**; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; OR (iv) possible changes in methods or levels of payments for benefits or services under those programs	☐
6	**Taste and food quality evaluation and consumer acceptance studies:** (i) if wholesome foods without additives are consumed, OR (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe by the Food and Drug Administration (FDA) or	☐

	approved by the Environmental Protection Agency (EPA) or the Food Safety and Inspection Services of the US Department of Agriculture	
	NONE OF THE ABOVE? – This study may not qualify for exemption.	☐

5. RESEARCH METHODS & INSTRUMENTS

Fill or paste with appropriate text in the editable spaces provided. The ""Review Questions" shown within closed boxes are locked and cannot be edited until a review has been completed.

5.1 Protocol Summary – Use this section to summarize the entire protocol using all the steps presented in this protocol. Provide a step-by-step account all of the procedures and interventions/interactions to be experienced by the participants starting from the recruitment till debriefing. Also include time and resource commitments to the participants. Use subtitles or separate steps using paragraphs.

Recruitment:

Participants will be recruited from the MTSU Department of Psychology's Sona system, MTSU Psychology classes, social media, and Amazon's MTurk.

- The students recruited through Sona will sign up for a time to participate in the study. They will be provided with an online consent form.
- Students from the Psychology classes (not part of Sona) will be able to sign up for a time to complete the study. They will be provided with an online consent form.
- Participants from MTurk will sign up for the study online. They will be provided with an online consent form.
- Participants from social media (non-compensated) will sign up for the study online. They will be provided with an online consent form.

Completing the Study:

Using an on-line survey platform, participants will be asked to electronically provide informed consent and then will be asked to complete a survey that will ask them about their attitudes, experiences, preferences, and feelings regarding a variety of topics and issues. Participants will be asked to answer demographic questions and questions about their familiarity with performance evaluations and behaviorally anchored rating scales (BARS).

Participants will then be randomly assigned to either the 3 example BARS condition (Condition A) or the 5 example BARS condition (Condition B). Regardless of the condition to which they are assigned, participants will then view a series of videos in which a University profressor provides a lecture on a topic. Participants will be asked at the end of each lecture to rate the professor's teaching using the BARS rating format provided. The rating scale for both the three anchor BARS and the five anchor BARS will be the same but the number of examples provided for each point on the scale will vary. Upon completion of the study, participants will be provided information indicating that they have completed the study.

The total time commitment required for participants will be between 45 and 60 minutes. Participants will only be required to have access to a computer and the internet.

5.2 Study Description – Describe this study using the outline provided below:

5.2.1 Purpose

This research study seeks to determine if defining more behavioral anchors (5 examples vs. 3 examples) on behaviorally anchored rating scales increases rater accuracy in ratings of performance evaulations. Additionally, the study will investigate if there are differences in accuracy of ratings provided by individuals who demonstrate a higher degree of conscientiousness.

5.2.2 Background

Behaviorally anchored rating scales (BARS) were first established in 1963 and have since been used for performance evaluations in many organizations. Many research studies have been conducted comparing BARS to other rating scales and found that BARS overall produce less error and more accurate results. Very little research has been conducted regarding how many anchors on such scales should be defined.

5.2.3 Rationale for Using Human Subjects

Human beings typically conduct performance evaluations, thus the onlly way to determine the accuracy of their ratings would be to study human beings and their behaviors. In order to investigate these behaviors one would have to study human subjects.

5.2.4 Study Design

Using an on-line survey platform, participants will be asked to view videos of University profress providing a lecture on various course topics. Participants will be asked to rate the profressor's teaching using BARS. Participants will be randomly assigned to rate the lectures on BARS with three anchors defined (three examples of performance) or to rate the lectures on BARS with five anchors defined (five examples of performance). The rating scale for both the three anchor BARS and the five anchor BARS will be the same but the number of examples provided for each point on the scale will vary.

5.2.5 Other Information not Included Above

Not applicable.

REVIEW QUESTION A: Is the purpose of this protocol and the associated procedures/interventions clearly described to make a rational decision?
Reviewer Comments:
Investigator Response:

5.3 Data Type – Check all those apply and provide additional information as directed

- ☐ Existing data *(complete 5.3.6)*
- ☐ Biospecimen *(complete 5.3.7)*
- ☐ Educational *(complete 5.3.1-5)*
- ☐ Social *(complete 5.3.1-5)*
- ☒ Behavioral *(complete 5.3.1-5)*
- ☐ Physical interventions
- ☐ Psychological interventions

THESE ARE DISQUALIFIED

- ☐ OTHER(s)

5.3.1 COVID-19 Risk Assessment – Select one of the following

- ☒ Virtual or online interaction with NO direct physical contact with the participant
- ☐ **Direct physical interaction with the participant: Complete Section 5.6**

5.3.2 Data Acquisition - Select all that apply

☒ ***5.3.2.1*** **Survey**[8]

- ☐ Paper Survey — *Submit survey script/topics*
- ☒ Online Web-based Survey[10] — *Insert Weblink for the survey*

Qualtrics Link(s):
https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_bNIXX3VwXPiV5BP

Visit https://mtsu.edu/irb/FAQ/OnlineDataCollection.php for more information

SONA: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_a9l4TOJKQ3ymo6y

MTurk: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_esMmrbKrdKzvydw

Social Media: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_eCX8YqL8iu2Ek4u

☐ Verbal Survey ☐ Telephone ☐ Email *Submit survey script/topics*

☐ Survey through Social Media *Submit screen shots*

☐ OTHER modes *Explain below and submit documents*

Explain:

☐ ***5.3.2.2*** **Interview**[8] *Submit interview script/topics*

☐ ***5.3.2.3*** **Observation**[9]

Explain and describe the instruments

☐ ***5.3.2.4*** **Focus Group(s)**[9]

Explain and describe the instruments:

☐ ***5.3.2.5*** **Other Modes Not Included Above**

Explain and describe the instruments

Description:

[8] *Attach a list of survey/interview questions with the application*

[9] *Describe the instruments to be used in the observational study or to be used during focus groups*

10 All of the investigators MUST *complete "Internet Based Research" module under CITI SBR course*

5.3.3 ***Provide a short description of what is collected in section 5.3.2 above*****:** Participants will be recruited from the MTSU Department of Psychology's Sona system, from other courses, and through other recruitment methods (social media, on-line recruitment platforms, e-mail, etc.). Using an on-line survey platform, participants will be asked to electronically provide informed consent and then will be asked to complete a survey about their attitudes, experiences, preferences, and feelings regarding a variety of topics and issues. Participants will be asked to answer demographic questions and questions about their familiarity with performance evaluations and behaviorally anchored rating scales.

5.3.4 ***Explain how the data described in 5.3.2 will be collected:*** Participants will then be randomly assigned to either the 3 example BARS condition (Condition A) or the 5 example BARS condition (Condition B). Regardless of the condition to which they are assigned, participants will view a series of videos in which a University profressor provides a lecture on a course topic. Participants will be asked at the end of each lecture to rate the professor's teaching using the BARS rating format provided. The rating scale for both the three anchor BARS and the five anchor BARS will be the same but the number of examples provided for each point on the scale will vary. Upon completion of the study, participants will be provided information indicating that they have completed the study.

5.3.5 ***Describe how the data collected from 5.3.2 will be analyzed*:** Data analysis will include- but shall not be limited to a repeated measures MANOVA.

5.3.6 ***Existing Data – OTHER THAN BIOSPECIMEN***

- *Definition:* "Existing Data" corresponds to the generalizable information generated or collected from living individuals using an approved IRB protocol. If the data were already collected without an IRB protocol, then IRB approval will not be granted.
- *Data Release:* If the existing data are not publicly available, a **Data Release Certification** may be needed from the original owner of the data in order to obtain IRB approval

5.3.7 ***Biospecimen collected through a previously approved IRB protocol***

REVIEW QUESTION B: Is the data acquisition, usage and analysis clearly explained?
Reviewer Comments:
Investigator Response:

5.4 Research Site(s) - Where will the research be conducted?

☒ MTSU – Department(s)/Building(s) MTSU Department of Psychology's Sona System
☐ Public Place(s)
☒ OTHER[12] social media, on-line recruitment platforms, e-mail, etc.

[12] *Permission letter(s) from non-MTSU organizations must be provided as a scanned PDF of a message written on an official letter head signed by an official from the organization who has such authority. Forwarded emails, text messages and other non-verifiable formats will NOT be accepted.*

From: Kelsey Byerly
Sent: Thursday, March 11, 2021 6:18 PM
To: irb_submissions <irb_submissions@mtsu.edu>
Cc: Mark Frame <Mark.Frame@mtsu.edu>
Subject: Re: 21-1131 2q Review Eyerly and Frame Exempt Submission

Hi Moses,

I have attached a copy of the survey for each type of distribution (e.g., SONA, MTurk, and Social Media). The survey is the same across all three types of distribution; the only difference is the compensation section located in the informed consent at the beginning of the survey.

That is correct, there will be three Qualtrics links. I have provided all three below.

SONA: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_a9l4TOJKQ3ymo6y
MTurk: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_esMmrbKrdKzvydw
Social Media: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_eCX8YqL8iu2Ek4u
**The SONA link was the original link provided in the IRB form.*

If you need any additional information, please feel free to reach out.

Thank you,

Kelsey Byerly
Master's Student
Industrial-Organizational Psychology
Middle Tennessee State University

From: irb_submissions <irb_submissions@mtsu.edu>
Date: Thursday, March 11, 2021 at 1:48 PM
To: Mark Frame <Mark.Frame@mtsu.edu>
Cc: Kelsey <krb7r@mtmail.mtsu.edu>
Subject: RE: 21-1131 2q Review Eyerly and Frame Exempt Submission

Kelsey,

I have completed the review of this request for IRB exemption. I currently do not have a copy of the actual survey that would be taken by the participants. Please provide a hard copy; it can be a PDF or a word document. If each type of participant would take a different survey, then please provide different copies. Also, there is only one Qualtrics link that is given for reviewhile 3 versions of informed consent template has been provided. Please confirm if there will be 3 Qualtrics links.

Approval is granted pending your response.

Sincerely,

Moses M. Prabu, Ph.D.
Compliance Officer
Middle Tennessee State University (PO BOX 124)
Tel: +1 615 494 8918
Email: Moses.Prabu@mtsu.edu

-sent via Outlook Web App

5.5 What are the risks for the participants? – Describe in detail how this proposed study presents no more than minimal risk[13] to the participants.

Participants will have the option to exit the survey at any time without any consequences. Thus, participating in this study will present no risk to participants that would be greater than the risks they would experience in their daily life. Because none of the information is

sensitive or provocative, there is no risk of harm or potential discomfort other than those expected in daily life.

Moreover, participants responses will be not be identifiable and no identifiable information will be linked to their responses.

[13] *"Minimal risk" describes the probability and magnitude of harm or potential discomfort anticipated in the research are not greater than those ordinarily encountered in daily life. Also note that research that involves more than minimal risk will disqualify this study from exemption.*

5.6 If the participants will have direct interactions with other participants or with the investigators, then complete this section to describe how this protocol will address the risk due to COVID-19

NONE

REVIEW QUESTION C: If risks are necessary, are they minimized to an extent such that the participants are only exposed to the same amount of risk they would experience in their normal life?
Reviewer Comments:
Investigator Response:

5.7 What are the benefits of this study?

5.7.1 ***To the field of science, society or common good:*** The findings of this study will provide information on how to best conduct performance evaluations. Many organizations are implementing BARS into their performance evaluation process, therefore information should be available on how to get the most accurate results.

5.7.2 ***To the participants:*** Participants will benefit from this study by becoming familiar with performance evaluations and how they are used to rate performance.

NOTE: Include only the benefits the participants may receive in the context of this research. They would not receive this benefit outside this study. Please enter "The participants will not have any direct benefits" otherwise.

REVIEW QUESTION D: Does this study result in benefits that outweigh the potential risks?
Reviewer Comments:
Investigator Response:

6. PARTICIPANT INFORMATION

6.1 Research Participant Recruitment – Describe how you will recruit the participants (recruitment materials MUST be submitted with this form), indicate whether the participants are 18 years of age or older, estimate the approximate number of research participants and describe inclusion/exclusion criteria used in selecting the participants.

6.1.1 Recruitment Tool(s) – Visit https://mtsu.edu/irb/FAQ/Recruitment.php

☒Flyer

☐Word of mouth[14] ☒Email[14] ☐Telephone[14] ☐Regular Mail[14] (Submit sample)

[14]Send the recruitment transcript as a separate file for IRB review. If contacting the participants by email or telephone or regular mail, explain how you originally obtained their contact information.

☐Web posting – Explain how the initial contact will be made

☒Social media – EXPLAIN how the initial contact will be made

(social media post)

We are conducting a research study to determine how to best design performance evaluations.

As part of our survey, you will be asked questions about yourself, attitudes, opinions, and feelings.

Then, you will view videos of University professors providing lectures on a course topic and be asked to rate the professors' performance.

We recommend completing this study using a PC or tablet, instead of a mobile device or smart phone. Participants MUST have video and audio capability on their computers and MUST be able to view and hear the videos being presented. Please complete the study in an environment which will allow you to hear and view video content.

If you would like to participate in our study, please click the link below

https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_bNIXX3VwXPiV5BP

If clicking the link does not work, please copy the information above and paste it into your internet browser's address bar.

☐OTHER

6.1.2 Describe the recruitment strategy including the recruitment steps to be followed using the recruitment tools stated above: The researchers will post information about the study on social media websites and via e-mail. The messages will contain information related to the nature of the study and the time commitment required for the study. This information will be identical to that included in section 6.2 of this document. For student participants, the survey link will be made available to faculty members with suggested language to use regarding the study. For students in a university's research pool, the study will be made available and advertised on the university's research pool sign up system (e.g. Sona website).

6.2 Participant Description – Complete this section for all types of research including analysis of existing data (if previously collected data are used, then describe the source from whom the data were originally collected).

6.2.1 Participants' Age	18 and older
6.2.2 Participant Description	All participants over the age of 18 will be included. This includes university students and adults outside of universities.
6.2.3 Sample Size	1000
6.2.4 Inclusion Criteria	Adult 18 years of age or older
6.2.5 Exclusion Criteria	Participants must meet the following criteria: *Be at least 18 years of age
6.2.6 Compensation*	Students completing the study for as part of a class will receive class credit for their participation in the study. Students completing the study as part of a university's research pool will receive credit per the research pool guidelines and requirements. Adults participating through Amazon's Mechanical Turk will receive monetrary compenseation for participants. Adults participating through other recruitment sources will not receive compensation *Complete Section 6.2.6.1 below*

All recruitment materials must be submitted for IRB approval, including transcripts of personal correspondences. If the participants are to be drawn from an institution or an organization that has the authority to allow its members to participate in human subject research, then proper approval notifications from that institution MUST be submitted with this application.

6.2.6.1: Compensation Documentation Requirement:

- Will the compensation have monetary value?
  ☒ Yes ☐ No
- If yes, will you be using MTSU funds or funds from an institution that requires documentation proof of how the compensation was disbursed?
  ☐ Yes ☒ No
- If No, then no further action is needed. Continue to 6.3 below.
- If you selected Yes, then additional documentation will be needed and the Compliance Office will direct you accordingly

6.3 Enrolling Participants from Psychology Research Pool (SONA): ☐ NOT Applicable

Complete this section ONLY if you plan to involve the students registered under this research pool. The information provided here will be relayed as it appears here to the student volunteers of the Psychology Research Pool.

6.3.1 Title: Rate the Professors

6.3.2 Abstract: Provide a short abstract (2-3 sentences)

The purpose of this study is to determine how to best design course evaluations. Participants will be asked a series of questions regarding their knowledge of performance evaluations. Participants will be asked to view and then rate Univeristy professors providing lectures on various course topics.

6,3.3 Description: Complete a short description of this project by elaborating the risks, benefits and other information necessary for the research pool volunteers to make an informed decision.

We recommend completing this study using a PC or tablet, instead of a mobile device or smart phone. Participants MUST have video and audio capability on their computers and MUST be able to view and hear the videos being presented. Please complete the study in an environment which will allow you to hear and view video content. Participants will benefit from this study by becoming familiar with performance evlaluations and how they are used to rate performance. Participation in this study will help develop research in the field of Industrial/Organizational Psychology.

*Visit **http://capone.mtsu.edu/wlangsto/ResearchPoolPage.html** for further information.*

6.4 Recruiting Amazon Mechanical Turk workers ☐ NOT Applicable

Complete MTurk Additional information Page Form F023 (https://mtsu.edu/irb/forms.php)

MTurk Additional Page is attached: ☐ NO (The protocol will not be reviewed) ☒ Yes

6.5 Enrolling Qualtrics Panel members as participants ☒ NOT Applicable

Complete Qualtrics Panel Additional information Page Form F023b from the IRB Forms page (https://mtsu.edu/irb/forms.php)

Qualtrics Panel Additional Page is attached: ☐ NO (will not be reviewed) ☐ Yes

REVIEW QUESTION E: Did you find the recruitment practice to be proper?
Reviewer Comments:
Investigator Response:

REVIEW QUESTION F: Does the proposed inducement sound reasonable without conflicts of interest or coercion?
Reviewer Comments:
Investigator Response:

6.6 Confidentiality – Describe in detail how you propose to protect the confidentiality of the information obtained from the participants. Mention if anyone outside the research team will have access to the participant information.

The data will be collected anonymously and deidentified.

6.7 Data Storage - Where will the data/records relating to the human participants be stored?

The data will be stored in Dr. Frame's office. The office has a locked door. The electronic files will be kept secure

Mandatory Data Storage Requirements:

- All Study related records (documentation of informed consent, surveys, study notes, data records, and all correspondence) be stored securely for **at least 3 years** after data collection ends.
- Additionally, the Tennessee State data retention requirement may apply (*refer MTSU Policy 129:* https://www.mtsu.edu/policies/general/129.php).

- Records must be stored securely in a faculty member's office on campus for 3 years. (Or another secure location if there is reason to believe the faculty member's office is not secure. These arrangements must be approved).
- Subsequently, the data may be destroyed in a manner that maintains confidentiality and anonymity of the research subjects.

REVIEW QUESTION G: Has/Have the researcher(s) done everything possible to protect the participants' anonymity and confidentiality?
Reviewer Comments:
Investigator Response:

8. INFORMED CONSENT

Investigators must remember that the consent process is like a conversation; it is not merely a document. Therefore, this process must be one of the center theme of your protocol in addition to protecting the autonomy and confidentiality of the subjects. The investigators are required to fully inform the participants on all of the activities to be carried out in the study and they must obtain consent from the latter prior to data collection. An informed consent document can be obtained from the MTSU IRB website. **Respond to these questions after completing the MTSU-approved informed consent template:**

8.1 Who will obtain informed consent? (Full Name(s))	Kelsey R. Byerly and Mark C. Frame
8.2 How will the consent be obtained? (Describe how consent will be administered and obtained)	Electronically
8.3 What language(s) is the text?	English
8.4 Where will the consent be obtained?	Consent will be obtained online

REVIEW QUESTION H: Is there enough evidence that the subjects are adequately informed and the autonomy of the participants respected?
Reviewer Comments:
Investigator Response:

REVIEW QUESTION I: Are the informed consent processes/documents fair and appropriate?
Reviewer Comments:
Investigator Response:

9. CITI TRAINING

This application WILL NOT be reviewed if the training for all of the investigators is incomplete

- The entire investigating team must complete "Social and Behavioral Research" basic training module
- Students must also complete "Students in Research" module in addition
- Study-specific and participant-specific modules/training must also be completed
- **Click here or visit http://www.mtsu.edu/irb/requirements.php to learn more**

The following CITI course(s) and modules are mandatory. Review your CITI training certificate and check boxes for all those modules that have been completed by the entire research team.

☒ Social & Behavioral Research (SBR)	
Modules for All Researchers	Modules required based on researcher status and the study
☒ Belmont Report and CITI … (ID: 1127) ☒ History and Ethical Principles - SBE (ID: 490) ☒ Defining Research ….. - SBE (ID: 491) ☒ The Federal Regulations - SBE (ID: 502) ☒ Assessing Risk - SBE (ID: 503) ☒ Informed Consent - SBE (ID: 504) ☒ Privacy and Confidentiality - SBE (ID: 505) ☒ Conflicts of Interest in …. (ID: 488) ☒ MTSU Module DEMO (ID 1073)	☒ Students in Research (ID 1321) **MANDATORY FOR STUDENTS** ☒ Research with Prisoners – SBE (ID: 506) ☒ Research with children – SBE (ID 507) ☒ Research in Public ….. Schools – SBE (ID 508) ☒ International Research – SBE (ID 509) ☐ International Studies (ID 971) ☒ Internet-based research – SBE (ID 510) ☒ Research and HIPAA …. (ID 14) ☐ Research on Workers/Employees (ID 483) ☐ Hot Topics (ID 487) ☐ IRB Member module (ID 816) ☐ IRB Administrators …. (ID 13813)

REVIEW QUESTION J: Are the reseachers' experience/qualification/training adequate?
Reviewer Comments:
Investigator Response:

10. ATTACHMENTS AND ENCLOSURES

Documents or Websites Included in this IRB submission:

☒ Informed Consent form ☐ Surveys/questioners/interview scripts

☒ Recruitment materials and transcripts ☐ Official Permission Letter(s)

☐ Prescreening/debriefing materials ☒ CITI certificates

☐ OTHER(S), Specify:

☒ Online link(s):

https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_bNIXX3VwXPiV5BP

- SONA: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_a9l4TOJKQ3ymo6y
- MTurk: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_esMmrbKrdKzvydw
- Social Media: https://mtsupsychology.az1.qualtrics.com/jfe/form/SV_eCX8YqL8iu2Ek4u

Separate the links by ";" for materials to be reviewed (video clips, literature data and etc.)

11. DECLARATION

PI Status:

☒ Student – Complete 11.1 and have faculty advisor/sponsor must fill 11.2
☒ Faculty/Staff – Complete 11.1 AND 11.2

11.1 Primary Investigator's Assurance

I, **Kelsey Byerly**, hereby certify that	Indicate acceptance by entering initials
1. As the PI of this study, I assure that this application packet has been fully completed by providing all essential and required information.	KRB
2. The information provided for this exemption request is accurate to the best of my knowledge.	KRB
3. All of the investigators have completed all research-specific CITI training; I will inform the IRB immediately if training deficiencies should occur.	KRB
4. Email addresses and contact information for all investigators are given.	KRB
5. Surveys, questionnaires, tests, interview forms etc. have been included.	KRB
6. Recruitment materials (OR/and) signup information for using Psychology research pool is completed *(Enter N/A if not applicable)*.	KRB
7. A filled informed consent form is attached.	KRB
8. PDF scan of all signed permission letters for researching at outside institutions (e.g., schools), is provided on official letterhead (Enter N/A if not applicable).	KRB
9. Once this protocol has been approved,	
• I will make every effort to protect the safety and welfare of the participants. I will inform the IRB immediately of any adverse events to the participants.	KRB
• Any deviations from the proposed methods will be reported immediately and changes will be implemented only after IRB approval.	KRB
• I will submit a status report of this study if directed by the IRB.	KRB
• I am aware of potential liabilities and sanctions for failure to adhere to my proposed protocol from IRB and non-IRB entities within MTSU and I agree to comply with those requirements.	KRB
• I assure that the data collected during this study and other records will be stored in a secure place within MTSU, such as the office of an MTSU faculty member. I also assure that the records will be stored for at least three years after the active data collection has been ceased.	KRB

PI[14] Kelsey R. Byerly Date: 03/02/2021

[14]Student PIs must complete this section using their MTSU FSA account

11.2 Faculty Investigator's Assurance

This section must be completed by an MTSU faculty member regardless if the PI is a student or not. An MTSU faculty member must read and endorse this section if the applicant is a student. Preferably use your MTSU FSA account when completing this section. If using a home computer, please ensure that you use a licensed version of MS Office for capturing the identity of the signee. Please visit the Faculty Information page http://www.mtsu.edu/irb/FAQ/Faculty.php before signing off this form.

I, **Mark C. Frame**, hereby certify that	Indicate acceptance by entering initials
1. This project will be carried out under my direct supervision	MCF
2. The investigators are competent and professional to work with human subjects and they comply with all of the provision required for the approval of this protocol	MCF
3. I have read this application thoroughly and I attest to its scientific merit.	MCF
4. I am fully aware of the activities to be performed under this exemption request.	MCF
5. All of the investigators, including myself, have completed all research-specific CITI training; I will inform training deficiencies to the IRB immediately.	MCF
6. Once this protocol has been approved,	
• I will report any significant or adverse events related to this study to the IRB within 72 hours of when I become aware of such incidents. I will also report breaches, such as, negligence or compromise to participant confidentiality or study-related injuries/discomforts to the participant.	MCF
• I take full responsibility to review any future changes or alterations to this study before a formal request is submitted to the IRB. Any deviations from the proposed methods will be reported immediately and changes will be implemented only after IRB approval	MCF
• I am aware of potential liabilities and sanctions for failure to adhere to my proposed protocol from IRB and non-IRB entities within MTSU and I agree to comply with those requirements[16]	MCF
• I assure that the data collected during this study and other records will be stored in a secure place in my Office or in my Department Office. I also assure that the records will be stored for at least three years after the active data collection has been ceased.	MCF
• I agree to meet with the investigators on a regular basis to monitor the study progress and compliance. I will retain records of such meetings, like email transactions and other verifiable communication records. I will also document specific conversations that would entail the welfare of the participants and other courses of actions	MCF

Faculty[15] Mark C. Frame Date: 03/04/2021

[15]Preferably complete this section using using your MTSU FSA account

[16]Faculty Sponsor Responsibilities - http://www.mtsu.edu/irb/FAQ/Faculty.php

INSTRUCTIONS FOR SUBMISSION:

- This application and support documents must be submitted by the faculty member who signed Section 11.2.
- Send all documents as separate files but in a single email to irb_submissions@mtsu.edu
- If multiple emails had to be sent due to memory insufficiency, then provide a proper explanation in each email
- Submit all IRB forms in their original MS Word format – DO NOT CONVERT TO PDF

The REVIEW STEPS

- The Office of Compliance (OC) will issue an IRB ID if the submission is determined to be complete
- If the application is incomplete, then the IRB request will be returned with no action
- Once the OC confirms that the application is complete, a reviewer will inspect the application packet and will enter any comments or request for additional information in the appropriate space provided within this form
- This form will be sent back to the investigators with reviewers' comments
- The investigators will receive any review comments, request for clarifications or recommended revisions along with other concerns. The review process is iterative and it depends on how swiftly the investigators are able to address all reviewers' concerns.
- Once a final approval has been issued, a "locked" version of this form will be sent to the investigators to be used as a guideline for their study.

12. REVIEWER SECTION
(Office Use Only)

Exempt Pre-Review Checklist	Y	N	N/A	Reviewer Comments
Application is complete	☐	☐	☐	
Informed consent is complete	☐	☐		
Recruitment/Debriefing is provided	☐	☐	☐	
Link for web-based research – TRAINING REQD	☐	☐	☐	
CITI Training Complete (PI, FA, Co-Investigators)	☐	☐	☐	
Application Appendices	☐	☐	☐	
Faculty Endorsement	☐	☐	☐	
Off-site Permission Letters	☐	☐	☐	
Research Instruments and Tools (i.e. Surveys)	☐	☐	☐	
Grant Information/Source of Funding Provided	☐	☐	☐	
Participant Pool	☐	☐	☐	
Sample Size	☐	☐	☐	
Restrictions	☐	☐	☐	

Exempt Designation Criteria	Y	N	Reviewer Comments
Subjects are considered "Vulnerable" according to OHRP's subpart definition [Examples – prisoners, cognitively impaired, seriously ill, pregnant women, minors (other than educational research)]	☐	☐	
Behavioral information collected in this study could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the individual's financial standing, employability or reputation	☐	☐	
Data involves sensitive information or personal aspects of the subject's behavior (drug/alcohol use, illegal conduct, sexual behavior, mental health, etc.)	☐	☐	
Except for researching normal education practices, will this study involve minors (under 18)?	☐	☐	
The subjects may be exposed to discomfort or stress beyond the levels encountered in daily life	☐	☐	
Video- or audiotaping is conducted	☐	☐	

13. IRB ACTION

Review Summary:	**Yes**	**No**
a. Is the purpose of this protocol clear?	☐	☐
b. Did you find the recruitment practice to be proper?	☐	☐
c. Does the proposed inducement sound reasonable?	☐	☐
d. Are the researchers' experience adequate?	☐	☐
e. Is there enough evidence that the subjects are adequately informed?	☐	☐
f. Are the informed consent process/documents appropriate?	☐	☐
g. Will the researchers protect the participants' confidentiality?	☐	☐
h. If risks are necessary, are the minimized to the maximum extent?	☐	☐
i. Does this study result in benefits that outweigh the potential risks?	☐	☐
j. Did the researcher(s) clearly explain the data usage?	☐	☐

If there is any reason why you may not be able to check "Yes" for all of the above questions, then please summarize your concern below:

Applicability:
Choose the criteria for IRB exemption: *(2) Educational Tests*

Correspondences - Enter review correspondences and paste email threads in the space below:

Recommendation:

Level of Risk:	☒ Lower than Minimal	☐ Greater than Minimal
Exemption Decision	☒ Exempt	☐ Revise and Resubmit
	☐ Defer (Expedited/Full)	☐ Not a "research"

Moses Prabu — 03/11/2021

(Reviewer's OC ID) — (Date of Determination)