

Fifty Years and Going Strong: What Makes Behaviorally Anchored Rating Scales So Perennial as an Appraisal Method?

Sukumar C. Debnath, D.B.A.

Associate Professor of Management
Department of Management & Marketing
Prairie View A&M University
Prairie View, TX 77446

B. Brian Lee, Ph.D., CPA

Professor of Accounting
Department of Accounting, Finance, & MIS
Prairie View A&M University
Prairie View, TX 77446

Sudhir Tandon, Ph.D.

Associate Professor of Marketing
Department of Management & Marketing
Prairie View A&M University
Prairie View, TX 77446

Abstract

Behaviorally anchored rating scales (BARS), compared to other appraisal methods, have the most potential to be an integral part of human resource management functions and contribute significantly to organizational success. It is not an accident that the BARS method has proved to be perennial and resilient in terms of survival over the last fifty years, and continues to sustain the interest of researchers and practitioners alike. Based on a synthesis of the existing literature, this paper explores the reasons underlying its resiliency. The paper provides a unique perspective by bringing together the more obvious factors and also unearthing rather implicit qualities, rarely emphasized elsewhere, which make BARS a perennial appraisal method. The implications of the findings (from motivational, equity, and utility perspectives) as well as challenges and future research needs are also discussed.

Keywords: Behaviorally Anchored Rating Scales, BARS, Performance Appraisal Method, Perennial Evaluation Method, Behavior Based Appraisal Method

1. Introduction

Performance appraisal is central to various human resource (HR) activities, such as recruitment, selection, training, development, and compensation (Landy, Farr, & Jacobs, 1982; Tziner, Joanis, & Murphy, 2000), and has always been considered to be an essential element of organizational success (Cintron & Flaniken, 2011). Studies provide support that performance appraisal, when conducted in an appropriate manner, can significantly influence both performance (Absar, Nimalathasan, & Mahmood, 2012; Al-Byadi, 2014) and effectiveness (Spinks, Wells, & Meche, 1999) of organizations. However, if appraisal is perceived to be unfair, it can negatively affect a whole range of organizationally desirable attitudes and behaviors, such as organizational commitment (Atteya, 2012; Abdulkadir, Isiaka, Adedoyin, 2012), turnover intention (Mustapha & Daud, 2012), job satisfaction, and performance (Atteya, 2012).

Currently, there are a host of performance appraisal methods, available to practitioners, which may be broadly classified into: (a) *trait methods*, such as graphic rating scales and forced choice method, (b) *results methods*, such as productivity and management by objectives, and (c) *behavioral methods*, such as behaviorally anchored rating scales and behavioral observation scales (Snell & Bohlander, 2012).

Many of these methods have faced criticisms and suffer from shortcomings that contribute to counterproductive organizational consequences (Elsbach & Cable, 2012; Jacobs, Kafry, & Zedec, 1980). While practitioners attempt to avoid certain appraisal methods (e.g., forced-choice rating system) and continue to use some others despite their shortcomings (Cardy & Dobbins, 1994), *the behaviorally anchored rating scales* (BARS) method continue to attract both researchers and practitioners alike since its introduction by Smith and Kendall (1963).

Interestingly, while graphic rating scales are the most widely used appraisal format in organizational settings, more studies have investigated BARS than any other methods (Cardy & Dobbins, 1994). As the literature reflects, research on BARS spiked in the 1970s with at least forty published studies, which was followed by a declining number of studies during the subsequent period of 1980s through 2010s. It must be emphasized that the decline in research on BARS was primarily due to the call for a moratorium on format research by Landy and Farr (1980), which essentially terminated appraisal format research in the coming years (Cardy & Dobbins, 1994). Another factor, contributing to this slowdown, could presumably be the sense of disappointment among researchers since research findings supporting the superiority of BARS over other appraisal methods were not equivocal on certain issues (see Bernardin, Morgan, & Winne, 1980; Jacobs et al., 1980). Nevertheless, the literature demonstrates that BARS has defied the call for such a moratorium, as is evident from the continuous streak of investigation involving BARS (as the citations throughout this paper indicate), though at a slower pace.

Furthermore, the versatility and irreplaceable nature of BARS are obvious from its application by researchers in areas that go well beyond typical performance evaluation. For example, BARS have been applied in the context of assessment centers (e.g., Schleicher, Day, Mayes, & Riggio, 2002), motivation (e.g., Landy & Guion, 1970), educational programs (e.g., Kavanagh & Duffy, 1978), experiential learning in college classrooms (e.g., Ohland, Layton, Loughry, & Yuhasz, 2005; McIntyre & Gilbert, 1994), structured interviews (Maurer, 2002), evaluation of team-member performance (Ohland et al., 2012), assessment of skills from web based training (Bussaman, Ruangsuwan, & Suksri-ngam, 2013), and morale (Motowidlo & Borman, 1977). The resiliency of BARS is also evident from its increased utilization by practitioners as well as organizations (Grote, 1996). For example, a survey of hotel industry by Woods, Sciarini, and Breiter (1998) revealed that out of the 389 participating hotels, 160 had used BARS. Therefore, BARS appear to be promising enough to sustain continued attention of both researchers and practitioners since its introduction fifty years ago.

The important question to ask is what makes BARS so resilient. The authors are not aware of any research that has investigated BARS from this perspective. The primary purpose of this paper is to explore the underlying factors, which make the BARS not only such a die-hard method, but also help it survive and prosper in the face of resistance. We offer a unique perspective involving BARS by evaluating inherent qualities of BARS in the light of relevant theories, unearthing the implicit factors rarely emphasized in the literature, and identifying motivational aspects of BARS, all of which are the likely reasons for its endurance.

As for the organization of the paper, a brief discussion of BARS is presented in the next section. Based on the existing literature, various underlying factors contributing to BARS' perennial characteristics are identified and presented. This is followed by a discussion related to implications, limitations, and future research needs involving BARS.

2. A Brief Overview of the BARS Method

Depending on the job under consideration, BARS may consist of a set of five to ten vertical scales. Each scale represents a major performance dimension of the job and is usually anchored by five or more critical incidents that reflect highly effective to highly ineffective observable job behaviors relevant to the job dimension under consideration (Snell & Bohlander, 2012). Scale values are assigned to the critical incidents, which correspond to the approximate degree of effectiveness with the highly effective behavior being assigned the highest value on the scale. The major performance dimensions for a job and the critical incidents for each dimension are identified through job analysis by future scale users (e.g., employees actually performing the job and their managers), who are expected to be thoroughly knowledgeable about the job (Bernardin & Beatty, 1984). In developing critical incidents, the emphasis is on incorporating job related behaviors that are observable and reflect various levels of desirable performance.

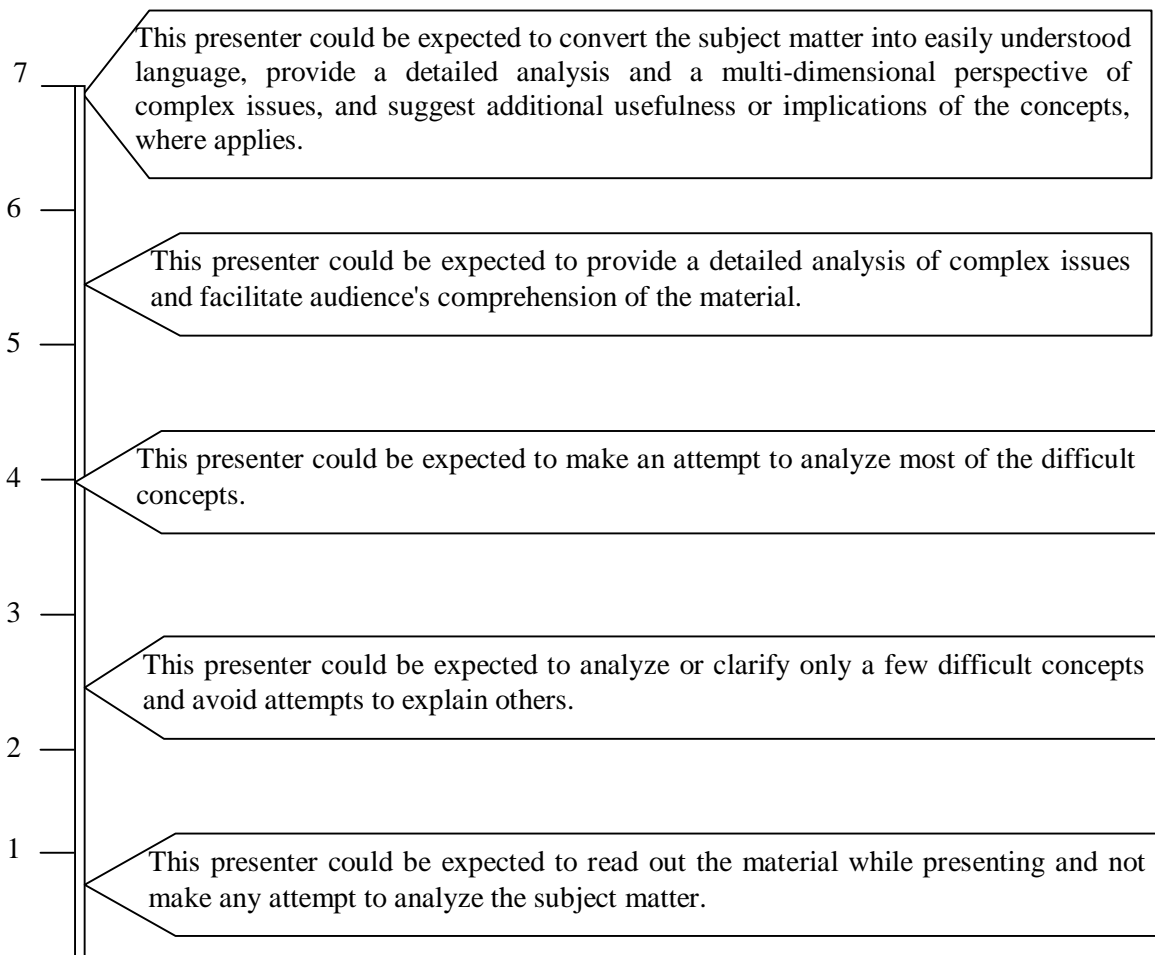
The BARS development process appears to be flexible which is evident from the fact that researchers have used several different procedures as well as scaling formats in developing BARS (Bernardin, LaShells, Smith, and Alvares, 1976b).

As mentioned earlier, the number of vertical scales (BARS) may vary from one job to another since it depends on the major performance dimensions of a job under consideration. The scale values (e.g., 1 to 5, 1 to 7, or 1 to 10) as well as the number of critical incidents anchored on a scale can also vary depending on the development procedure and appropriateness of the situation. A final version of BARS should be a jargon-free instrument that is closely related to the requirements of a given job (Ivancevich, 1992).

For example, the BARS presented by Bernardin (1977) for evaluating college professors was developed jointly by the instructor and the students and consisted of the following seven major performance dimensions: *testing procedures, student-teacher relations, organizational skills, communication skills, instructor knowledge, subject relevance, and utility of assignments*. Similarly, BARS related to classroom presentations by college students may involve several dimensions including *organization of the presentation, basic knowledge of the presentation material, analysis of the subject matter, and communication skills*. Table 1 provides an example of BARS related to the dimension, 'analysis of the subject matter'.

Table 1: BARS for Analysis of the Subject Matter (Under Student Presentation)

This area of presentation concerns the presenter's ability to breakdown and simplify the subject matter including complex issues, so that the audience can easily understand and follow the content being presented. Based on a thorough preparation, the presenter should be able to provide detailed analyses and multiple perspectives on issues, where relevant, and suggest real-life implications of the concepts.



3. What Makes BARS Resilient?

As discussed earlier, the perennial nature of the BARS method is evident through its survival, utilization, and continued exploration by both practitioners and researchers over the last few decades.

Defying the call for a moratorium related to format research by Landy and Farr (1980), numerous studies have explored BARS (Hom, DeNisi, Kinicki, & Bannister, 1982; Jacobs et al., 1980), and the interest in the research and application of BARS remains current (e.g., Bussaman et al., 2013; Catano, Darr, & Campbell, 2007; French, Goodman, & Morrison, 2014; Hauenstein, Brown, & Sinclair, 2010; London, Mone, & Scott, 2004; MacMillan, Entin, Morley, & Bennett Jr., 2013; Millán & Navas, 2013; Ohland et al., 2012; Schraeder, Becton, & Portis, 2007; Tziner et al., 2000). Based on an analysis of BARS and review of the related literature, we present possible explanations for the resiliency of BARS.

3.1 The Built-in Advantage of BARS

A comparison of BARS with other appraisal methods, such as trait, results, and behavioral categories should make apparent the inherent advantages of BARS over these methods. To begin with, the characteristics as well as procedures used in developing the instrument render the BARS as a superior appraisal method as alluded to earlier. The process utilized in developing BARS tends to ensure that (a) BARS cover the entire performance domain of a job and include all major job dimensions based on an extensive job analysis (Jacobs et al., 1980), (b) critical incidents or descriptive anchors are developed to incorporate a range of highly effective to highly ineffective observable behaviors for each dimension, and (c) BARS is based on job-relatedness (Hom et al., 1982). This process contributes to the *face validity* of BARS (MacDonald & Sulsky, 2009). The development process and the structure of BARS both can be instrumental to: identifying specific behavioral strengths and weaknesses of an employee, facilitating performance improvement, selecting candidates for promotion, validating selection criteria, justifying disciplinary actions, and developing training programs (Jacobs et al., 1980).

On the contrary, other appraisal methods have inherent shortcomings. For example, Elsbach & Cable (2012) confirmed that managers using trait methods are significantly more likely to attribute traits (e.g., dependable, responsible, committed, or dedicated) to employees simply based on the extent of their face time at workplace, irrespective of the level of their performance. According to these researchers, trait-based evaluations are flawed and should not be used since they are not linked to employee contributions and do not help employees understand what to improve. Next, while results methods allow for evaluation of employees based on objective or tangible data (*hard criteria*), such as sales figures, certain job behaviors (e.g., cooperation), which cannot be easily observed and assessed (*soft criteria*), remain excluded from evaluations in spite of their significant influence on the outcomes. The Bureau of National Affairs (1975) survey estimated that objective indices are available for only 14 percent of the situations useful for performance assessment, including those that may have a weak link to performance (Cardy & Krzystofiak, 1991). Given this scenario, it will be difficult for HR specialists to use results methods in most jobs to perform evaluations meaningful enough for a managerial decision-making purpose.

Even among the behavioral methods, which specifically describe desirable and undesirable behaviors on the job, BARS stands out based on utility value. For example: the *critical incident* method restricts evaluations to only unusually superior or inferior performance of an employee; the *behavioral checklist* method limits evaluation to a predetermined list of behaviors; and *behavior observation scales* similarly provide a select list of behaviors so that evaluators can record the frequency of exhibited behaviors for the listed items. Therefore, these methods afford incomplete evaluations of performance behaviors as the rater has no discretion.

To sum up, one apparent reason for BARS resiliency lies in its built-in advantage over other subjective methods (Millard, Luthans, & Otteman, 1976; Schraeder et al., 2007) as well as objective methods which are limited in scope.

3.2 Ownership Stake of Raters and Ratees

The BARS instrument is developed by the actual users--appraisers and appraisees themselves (Bernardin & Beatty, 1984; McIntyre & Gilbert, 1994). In contrast, most appraisal instruments are usually developed by HR personnel or people who may not be involved in the actual evaluation process. The involvement in the BARS' development process is likely to promote a sense of satisfaction, feelings of ownership of BARS, and acceptance of the appraisal instrument among the users (Locke & Schweiger, 1979; Northcraft & Neale, 1990). The participation process is also expected to provide better role clarity to the employees (Baron, 1990; Dobbins, Cardy, & Platz-Vieno, 1990) and make appraisal more effective (Qureshi & Hassan, 2013).

3.3 Accommodation of both Hard and Soft Criteria in BARS

As mentioned earlier, the Bureau of National Affairs survey (1975) indicated that objective indices (hard criteria) are not always available for most job situations. In lieu of the missing hard criteria, soft criteria (e.g., supervisory impression) become central to most job evaluation instruments. Compared to many other appraisal instruments, the BARS method is viewed as the more appropriate method for evaluation of jobs with both hard and soft criteria (Jacobs et al., 1980). The development process of BARS is expected to capture the entire range of both hard and soft criteria for a position, and the incumbent supervisors and employees should be able to develop critical incidents or descriptive anchors related to these criteria based on their knowledge and experience.

3.4 Utilization and Qualitative Superiority of BARS

Researchers (e.g., Borman, 1986; Jacobs et al., 1980) indicated that the BARS instrument is clearly superior to other evaluation instruments when compared in terms of *utilization* and *qualitative criteria*, such as rating accuracy and feedback potential. For example, the design and application procedures of BARS promote rating accuracy as it allows raters the freedom to indicate true performance of ratees; on the contrary, most other appraisal methods tend to embed mechanisms to maneuver raters into giving inflated evaluations (Schmitt & Klimoski, 1991). From the utilization perspective, a review of the existing research by Cardy and Dobbins (1994) revealed that the extensive feedback based on BARS--as opposed to limited feedback from graphic rating scales or forced-choice scales--led to a higher level of ratee satisfaction with BARS compared to the competing methods. Feedback based on BARS was more acceptable to the recipients (Krein, 1990) as well as more effective in bringing about behavioral changes among them as compared to feedback from alternative evaluation formats (Hom et al., 1982). In addition, as prior studies (e.g., Murphy & Cleveland, 1995) indicate, even employees who are not personally involved in the development of BARS also tend to view the scales favorably because of their reliance on colleagues' feedback. Moreover, the collectivist culture stemming from the development process of BARS tends to offer an added advantage to raters with regard to the appraisal process (MacDonald & Sulsky, 2009).

3.5 Common Frame-of-Reference

Research findings indicate that the practice of involving raters in developing BARS provides *frame-of-reference training* to the raters (Woehr & Huffcutt, 1994). Besides the training, BARS presents all raters with a *common frame-of-reference* (in the form of critical incidents) during evaluation. The common frame-of-reference should help improve inter-rater reliability (Cardy & Dobbins, 1994) and contribute to higher rating accuracy with BARS than other appraisal methods (Bernardin & Beatty, 1984). Studies (e.g., Campbell et al., 1973; Millard et al., 1976) have also indicated lower rater errors related to central tendency, leniency, and halo errors with BARS compared to graphic or summated rating scales.

3.6 Rater Discretion and Flexibility of BARS

Many evaluation instruments (e.g., graphic rating scales, forced choice method, behavioral checklist method, and behavioral observation scales) are inflexible and also restrict or force raters to choose from a limited number of options, even though the choice may not reflect the true performance. The critical incidents--arranged in order of effectiveness along with the corresponding scale values --for each performance dimension in BARS are meant to guide the rater in choosing an appropriate point on the scale continuum that reflects the true performance of the ratee (Bernardin, et al., 1976b; Bernardin & Smith, 1981; Cardy & Dobbins, 1994; Smith & Kendall, 1963). The freedom allowed to raters and the flexibility of BARS for rating purpose should contribute to rater motivation and rating accuracy.

3.7 Perception of Fairness

The involvement of employees in the BARS development process should instill a strong sense of fairness among the employees. When examined in the light of the *organizational justice theory* (Colquitt, 2001; Greenberg & Colquitt, 2004; Thurston Jr & McNall, 2010), the BARS development process is expected to enhance employees' fairness perception related to: (a) *distributive justice* that evaluations are designed to capture the actual performance behavior, (b) *procedural justice* that the process ensures that employee inputs are solicited prior to evaluation, and (c) *informational justice* that each step in the process (such as, clarification of performance behaviors and standards, performance related feedback, and explanation of decisions) is transparent.

The perceptions of fairness can promote satisfaction, motivation, and desirable behaviors among employees (Atteya, 2012; Thurston Jr & McNall, 2010). Research has confirmed that the BARS development process and perceptions of distributive justice can lead to employee satisfaction with the appraisal systems (Thurston Jr & McNall, 2010). On the contrary, many of the other appraisal methods (e.g., graphic rating scales, mixed standard scales, or force-choice method) may be perceived as unfair because of their shortcomings related to distributive, procedural, or informational justice (Cardy & Dobbins, 1994). Therefore, fairness perceptions can play a big role in supporting the resiliency of BARS.

3.8 Legal Defensibility

The Uniform Guidelines on Employee Selection Procedures and various regulations related to equal employment--such as *Equal Pay Act of 1963*, *Civil Rights Act of 1964 and 1991*, *Fair Employment Practices Laws*--prohibit discrimination in various HR activities including pay, promotion, retention, demotion, and dismissal based on race, color, religion, sex, or national origin. An employer charged with discrimination must demonstrate that employment practices are job related as well as consistent with business practices and that the evaluation procedure is valid (Snell & Bohlander, 2012).

The criteria identified by Malos (1998) for a legally sound performance appraisal system include, among other things, that an evaluation should be: (a) job-related, (b) based on behaviors, not traits, (c) related to specific functions, and (d) free from global assessments. Hauenstein et al. (2010) observed that the BARS method is very much aligned with the criteria recommended by Malos. Therefore, it is at a more advantageous position to meet legal requirements compared to many other methods, and can help save employers from expenses triggered by employment discrimination lawsuits and the embarrassments that follow them (Bushardt, Fowler Jr. & Debnath, 1985).

3.9 Supportive Findings Related to the Psychometric Characteristics of BARS

Literature indicates that BARS has been investigated by numerous researchers almost on a continuous basis over the last few decades. Kingstrom and Bass (1981), based on their review of a number of empirical investigations, found that BARS exhibited significant convergent as well as discriminant validity across the studies examined. In some cases, BARS showed greater convergent and discriminant validity compared to alternative scales (e.g., Arvey & Hoyle, 1974; Campbell, Dunnette, Arvey, & Hellervik, 1973). More recently, the results of several other studies (e.g., Harrell & Wright, 1990; MacMillan, et al., 2013; Millán & Navas, 2013; Ohland et al., 2012) provided further support for both reliability and validity (including content, construct, convergent, and predictive) of BARS.

With regards to rater errors (e.g., leniency and halo errors) and inter-rater reliability, previous studies indicated that the BARS method was either as good as (e.g., Bernardin, 1977; Borman & Vallon, 1974; Burnaska & Hollmann, 1974) or superior to (e.g., Campbell et al. 1973; Borman & Dunnette, 1975; Ohland et al., 2005) several other methods. Another important research finding involving BARS is that the rating bias from one performance dimension of BARS does not permeate the other dimensions (Murphy & Pardaffy, 1989) and thus BARS can limit the rater errors from spreading into multiple dimensions. The studies involving rating accuracy, which are very few to date (Cardy & Dobbins, 1994), reflect that the BARS is either as accurate as (e.g., Borman, 1979) or superior to (e.g., Benson, Buckley, & Hall, 1988) several other methods. Therefore, based on the findings of prior studies, the level of reliability, validity, and rating accuracy of BARS may be considered reasonably good.

4. Discussion

It is not an accident that the BARS method has proved to be perennial and resilient in terms of survival over the last fifty years, and continues to garner interest of researchers and practitioners alike. Based on a synthesis of the existing literature, this paper provides a unique perspective regarding its resiliency by bringing together the more obvious factors, and also unearthing rather implicit qualities of BARS rarely emphasized in the previous studies. The literature (e.g., Hauenstein et al., 2010) on *best practices* in performance appraisal continues to suggest the use of a behavioral anchor format, particularly the BARS method, and the discussion involving its resiliency should provide some insights into why it is recommended. However, there are several issues which need to be addressed if this method is to do well in the future, such as its limitations, challenges, and an existing research gap.

4.1 Practical Implications

The analysis presented in this paper reveals or reinforces several important aspects of the BARS method, which sets itself apart from most other appraisal methods. The development and application procedures of BARS allow for participation and empowerment of employees, and also contribute to their fulfillment of needs related to self-esteem, recognition, achievement, discretion, and control. Combined with the perception of equity, these attributes can instill a strong undercurrent of motivation into raters and ratees (Griffin & Moorhead, 2012). As reflected in the literature, another unique strength of BARS is the detailed and effective feedback potential, which should lead to performance improvement because of its clarity and acceptance by the employees. Finally, the BARS method is designed to include all aspects of a job, and is well positioned, compared to other methods, to meet legal requirements (Jacobs et al., 1980), and consequently, reduce litigation possibilities and the associated costs (Bushardt et al., 1985).

Therefore, while BARS continue to be a work-in-progress method requiring further research to resolve issues at hand (described in the next section), the analysis in this paper should make a strong case for exploiting the tremendous appraisal potential as well as utilization and qualitative values the method offers through its application in organizations.

4.2 Limitations, Challenges, and Future Research Needs

The literature demonstrates that researchers have used several different development procedures and scaling formats for BARS at times affecting the results related to inter-rater reliability and rater errors (Bernardin, Alvares, & Cranny, 1976a); this may limit our ability to compare findings across various investigations and make inferences. Another major limitation of BARS is the high cost associated with the employee time and effort required to develop and use the instrument (Borman & Dunnette, 1975; Jacobs et al., 1980). Bernardin and Smith (1981) argued that the time invested is well worth it because of the need to document and justify ratings. BARS developed, however, for a particular job may no longer be valid if the content and requirements of the job change subsequently, which may be a likely scenario because of the rapidly changing business environment. For example, employers are increasingly using remote working arrangements to accomplish goals (e.g., virtual office or telecommuting), which reduce the need for face time or physical presence at workplace. Incorporating non-traditional methods to accomplish a job that was performed earlier only on-site would necessitate a modification of the previously developed BARS to capture the new dimensions.

The BARS development process itself may present challenges for those involved in the process. Participants, while developing BARS, may easily agree on critical incidents which define highly effective or highly ineffective behaviors (since these are more salient), but agreement is difficult to achieve on the incidents reflecting mid-range or average performance (Grote, 1996; Hauenstein et al., 2010). Next, innovation--an important aspect of organizational life--may at times be difficult to capture in the critical incidents. Moreover, these incidents may serve to mold employee behaviors and may limit or counter innovations if employees do not think outside of the box.

Regarding evaluation process, researchers recommend that appraisers systematically reflect on an employee's behaviors over the period of evaluation and document them on the relevant BARS for subsequent use (Bernardin & Smith, 1981; Schmitt & Klimosky, 1991). However, in reality, the BARS instrument is frequently completed only at the end of the evaluation period, which could lead to rater errors.

Therefore, future research may be geared to address these aforementioned challenges in order to make the BARS method more effective and maximize the utilization value in organizational context. Researchers have called for additional investigations to further verify the relative rating accuracy and qualitative criteria of BARS compared to other rating methods (Cardy & Dobbins, 1994). In addition, while some researchers (e.g., Bernardin et al., 1976a) have made recommendations in an effort to standardize development procedures for BARS, further research seems to be necessary to establish concrete procedures for developing BARS aimed at achieving both parsimony (in reducing time and effort required for its development) and effectiveness.

Research should also be conducted to explore various other issues, such as: (a) the minimum number (or percentage) of employees to be involved in developing BARS to save employee time and cost without sacrificing the quality of the instrument; (b) how innovation possibilities can be captured in BARS, while maintaining the strengths of BARS; (c) whether creation of a *quality circle group*, to periodically review BARS in relation to the job, can improve the instrument and maintain its relevancy over time; and (d) how to minimize the difficulty encountered in anchoring the middle range of BARS, an issue which is attracting researchers lately (e.g., Hauenstein et al., 2010).

Therefore, further investigations are needed in a number of areas to make BARS more effective in terms of development, cost, and performance. Despite the shortcoming and challenges, researchers across the board appear to be optimistic regarding BARS and almost invariably agree that the relative value of BARS compared to other appraisal methods is not yet fully known (e.g., Jacobs et al., 1980). The discussion in this paper regarding the reasons for the resiliency of BARS reinforces the assertion by researchers that BARS is likely to endure and thrive in the future, provided some of the issues mentioned earlier are taken care of.

References

- Abdulkadir, D. S., Isiaka, S. B., & Adedoyin, S. I. (2012). Effects of strategic performance appraisal, career planning and employee participation on organizational commitment: An empirical study. *International Business Research*, 5(4), 124-133.
- Absar, N., Nimalathasan, B., & Mahmood, M. (2012). HRM-market performance relationship: evidence from Bangladeshi organizations. *South Asian Journal of Global Business Research*, 1(2), 238-255.
- Al-Byadi, A. K. (2014). The Relationship Between Performance Appraisals and Development at the Level Of Individual and Organization Case Study, Jordan Investment Board. *Interdisciplinary Journal of Contemporary Research in Business*, 5(10), 34-61.
- Arvey, R. D., & Hoyle, J. C. (1974). A Guttman approach to the development of behaviorally based rating scales for systems analysts and programmer/analysts. *Journal of Applied Psychology*, 59(1), 61-68.
- Atteya, N. M. (2012). Testing the Impact of the Human Resource Management Practices on Job Performance: An Empirical Study in the Egyptian Joint Venture Petroleum Companies. *International Journal of Business and Social Science*, 3(9), 105-119.
- Baron, R. A. (1990). Countering the effects of destructive criticism: The relative efficacy of four interventions. *Journal of Applied Psychology*, 75(3), 235-245.
- Benson, P. G., Buckley, M. R., & Hall, S. (1988). The impact of rating scale format on rater accuracy: an evaluation of the mixed standard scale. *Journal of Management*, 14(3), 415-423.
- Bernardin, H. J. (1977). Behavioral expectation scales versus summated scales: A fairer comparison. *Journal of Applied Psychology*, 62(4), 422-427.
- Bernardin, H. J., Alvares, K. M., & Cranny, C. J. (1976a). A recomparison of behavioral expectation scales to summated scales. *Journal of Applied Psychology*, 61(5), 564-570.
- Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston: Kent Publishing Company.
- Bernardin, H. J., LaShells, M. B., Smith, P. C., & Alvares, K. M. (1976b). Behavioral expectation scales: Effects of developmental procedures and formats. *Journal of Applied Psychology*, 61(1), 75-79.
- Bernardin, H. J., Morgan, B. B., & Winne, P. S. (1980). The design of a personnel evaluation system for police officers. *JSAS Catalog of Selected Documents in Psychology*, 10, 1-280.
- 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. (1986). Behavior-based rating scales. R. A. Berk, ed., *Performance Assessment: Methods and Applications*. Baltimore, MD: The Johns Hopkins University Press.
- Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, 64(4), 410-421.
- Borman, W. C., & Dunnette, M. D. (1975). Behavior-based versus trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology*, 60(5), 561-565.
- Borman, W. C., & Vallon, W. R. (1974). A view of what can happen when behavioral expectation scales are developed in one setting and used in another. *Journal of Applied Psychology*, 59(2), 197-201.
- Bureau of National Affairs. (1975). *Employee performance: Evaluation and control*. Personnel policies forum. Survey No. 108. Washington, DC: Bureau of National Affairs.

- 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.
- Bushardt, S., Fowler Jr, A., & Debnath, S. (1985). Behaviorally anchored rating scales: A valuable aid to the personnel functions. *Akron Business and Economic Review*, 16(2), 26-32.
- Bussaman, S., Ruangsuwan, C., & Suksri-ngam, P. (2013). The Development of Web Based Training System for Developing Core Competencies of Personnel at Rajabhat Maha Sarakham University. *The Journal of Human Resource and Adult Learning*, 9(1), 1-11.
- 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), 15- 22.
- Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: Alternative perspectives*. Mason, OH: South-Western Publishing Company.
- Cardy, R. L., & Krzystofiak, F. J. (1991). Interfacing high technology operations with blue collar workers: selection and appraisal in a computerized manufacturing setting. *The Journal of High Technology Management Research*, 2(2), 193-210.
- Catano, V. M., Darr, W., & Campbell, C. A. (2007). Performance appraisal of behavior-based competencies: A reliable and valid procedure. *Personnel Psychology*, 60(1), 201-230.
- Cintrón, R., & Flaniken, F. (2011). Performance Appraisal: A supervision or leadership tool. *International Journal of Business and Social Science*, 2, 29-37.
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of applied psychology*, 86(3), 386-400.
- Dobbins, G. H., Cardy, R. L., & Platz-Vieno, S. J. (1990). A contingency approach to appraisal satisfaction: An initial investigation of the joint effects of organizational variables and appraisal characteristics. *Journal of Management*, 16(3), 619-632.
- Elsbach, K., & Cable, D. (2012). Why showing your face at work matters. *MIT Sloan Management Review*, 53(4), 10-12.
- French, P. E., Goodman, D., & Morrison, K. C. (2014). An empirical evaluation of the influence of descriptive representation on human resource practice at the local government level. *Journal of Public Management and Social Policy*, 20(1), 47-66.
- Greenberg, J. E., & Colquitt, J. A. (2004). *Handbook of organizational justice*. Lawrence Erlbaum Associates Publishers.
- Griffin, R. W., & Moorhead, G. (2012). *Organizational Behavior: Managing People and Organizations*. Mason, OH: South-Western.
- Grote, D. (1996). *The complete guide to performance appraisal*. New York: American management Association.
- 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.
- Hom, P. W., DeNisi, A. S., Kinicki, A. J., & Bannister, B. D. (1982). Effectiveness of performance feedback from behaviorally anchored rating scales. *Journal of Applied Psychology*, 67(5), 568-576.
- Ivancevich, J. M. (1992). *Human Resource Management: Foundations of Personnel*. Boston: Richard D. Irwin.
- Jacobs, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. *Personnel Psychology*, 33, 595-640.
- Kavanagh, M. J., & Duffy, J. F. (1978). An extension and field test of the retranslation method for developing rating scales. *Personnel Psychology*, 31(3), 461-470.
- 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.
- Krein, T. J. (1990). Performance Reviews that Rate an 'A'. *Personnel*, 67(5), 38-41.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87(1), 72-107.
- Landy, F. J., Farr, J. L., & Jacobs, R. R. (1982). Utility concepts in performance measurement. *Organizational Behavior and Human Performance*, 30(1), 15-40.
- Landy, F. J., & Guion, R. M. (1970). Development of scales for the measurement of work motivation. *Organizational Behavior and Human Performance*, 5(1), 95-103.
- Locke, E. A., & Schweiger, D. M. (1979). Participation in decision-making: One more look. *Research in organizational behavior*, 1(10), 265-339.
- London, M., Mone, E. M., & Scott, J. C. (2004). Performance management and assessment: Methods for improved rater accuracy and employee goal setting. *Human Resource Management*, 43(4), 319-336.

- MacDonald, H. A., & Sulsky, L. M. (2009). Rating formats and rater training redux: A context-specific approach for enhancing the effectiveness of performance management. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 41(4), 227.
- MacMillan, J., Entin, E. B., Morley, R., & Bennett Jr, W. (2013). Measuring team performance in complex and dynamic military environments: The SPOTLITE method. *Military Psychology*, 25(3), 266-279.
- Malos, S. B. (1998). Current legal issues in performance appraisal. *Performance appraisal: State of the art in practice*, 49-94.
- Maurer, S. D. (2002). A practitioner-based analysis of interviewer job expertise and scale format as contextual factors in situational interviews. *Personnel Psychology*, 55(2), 307-327.
- McIntyre, F. S., & Gilbert, F. W. (1994). Improving performance in case courses: An argument for behaviorally anchored rating scales. *Marketing Education Review*, 4(1), 51-58.
- Millán, J. M., & Navas, M. F. (2013). Development of an evaluation performance scale for social educators in child protection centers. *Intangible Capital*, 9(3), 571-589.
- Millard, C. W., Luthans, F., & Ottemann, R. L. (1976). A new breakthrough for performance appraisal. *Business Horizons*, 19(4), 66-73.
- Motowidlo, S. J., & Borman, W. C. (1977). Behaviorally anchored scales for measuring morale in military units. *Journal of Applied Psychology*, 62(2), 177-183.
- Murphy, K. R., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. SAGE Publications, Incorporated.
- Murphy, K. R., & Pardaffy, V. A. (1989). Bias in behaviorally anchored rating scales: Global or scale-specific?. *Journal of applied psychology*, 74(2), 343-346.
- Mustapha, M., & Daud, N. (2012). Perceived performance appraisal effectiveness, career commitment and turnover intention of knowledge workers. *International Journal of Business and Social Science*, 3(19), 157-165.
- Northcraft, G. B., & Neale, M. A. (1990). *Organizational behavior: A management challenge*. Oak Brook, IL: Dryden Press.
- Ohland, M. W., Layton, R. A., Loughry, M. L., & Yuhasz, A. G. (2005). Effects of behavioral anchors on peer evaluation reliability. *Journal of Engineering Education*, 94(3), 319-326.
- Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., Layton, R. A., Pomeranz, H. R., & Schmucker, D. G. (2012). The Comprehensive Assessment of Team Member Effectiveness: Development of a Behaviorally Anchored Rating Scale for Self-and Peer Evaluation. *Academy of Management Learning & Education*, 11(4), 609-630.
- Qureshi, A., & Hassan, M. (2013). Impact of performance management on the organisational performance: An analytical investigation of the business model of McDonalds. *International Journal of Academic Research in Economics and Management Sciences*, 2(5), 54-76.
- Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology*, 87(4), 735-746.
- Schmitt, N., & Klimoski, R. J. (1991). *Research methods in human resources management*. Mason, OH: South-Western Publishing Company.
- Schraeder, M., Becton, J. B., & Portis, R. (2007). A Critical Examination of Performance Appraisals: An Organization's Friend or Foe?. *The Journal for Quality & Participation*, 30(1), 20-25.
- 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(2), 149-155.
- Snell, S., & Bohlander, G. (2012). *Managing human resources*. Cengage Learning.
- Spinks, N., Wells, B., & Meche, M. (1999). Appraising the appraisals: computerized performance appraisal systems. *Career Development International*, 4(2), 94-100.
- Thurston Jr, P. W., & McNall, L. (2010). Justice perceptions of performance appraisal practices. *Journal of Managerial Psychology*, 25(3), 201-228.
- Tziner, A., Joanis, C., & Murphy, K. R. (2000). A comparison of three methods of performance appraisal with regard to goal properties, goal perception, and ratee satisfaction. *Group and Organization Management*, 25(2), 175-190.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67(3), 189-205.
- Woods, R. H., Sciarini, M., & Breiter, D. (1998). Performance Appraisals in Hotels Widespread and Valuable. *Cornell Hotel and Restaurant Administration Quarterly*, 39(2), 25-29.