

Trends in Training

Trends in Training

Edited by

Brian J. O'Leary, Bart L. Weathington,
Christopher J. L. Cunningham
and Michael D. Biderman

Cambridge
Scholars
Publishing



Trends in Training

Edited by Brian J. O'Leary, Bart L. Weathington,
Christopher J. L. Cunningham and Michael D. Biderman

This book first published 2015

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2015 by Brian J. O'Leary, Bart L. Weathington,
Christopher J. L. Cunningham, Michael D. Biderman and contributors

All rights for this book reserved. No part of this book may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording or otherwise, without
the prior permission of the copyright owner.

ISBN (10): 1-4438-7825-1

ISBN (13): 978-1-4438-7825-8

TABLE OF CONTENTS

Preface	vii
Chapter One.....	1
Current Trends in Rater Training: A Survey of Rater Training Programs in American Organizations C. Allen Gorman, John P. Meriac, Joshua L. Ray, and Thomas W. Roddy	
Chapter Two	25
Using a Training Model to Improve Leadership Coaching Mark C. Frame and Emily K. Sanders	
Chapter Three	37
Examining the Efficacy of Outdoor Adventure Training Jennifer (Hopkins) Smallwood and Christopher J. L. Cunningham	
Chapter Four.....	61
Stress and Learning Performance Traci L. Carico and Bart L. Weathington	
Chapter Five	85
Leadership Training for Women Kathryn Thompson and Stephanie Hale	
Chapter Six.....	95
Massive Open Online Courses: The Future of Training or MOOC Ado about Nothing? Brian J. O’Leary and Tracye L. Poole	
Contributors.....	111

PREFACE

Organizations around the world spend millions of dollars each year on employee training. However, when economic downturns occur expenses for training are often the first to be cut. This is an unfortunate tendency because time and again it has been shown that organizations that invest in employees emerge from economic downturns as leaders and innovators.

To implement successful training, it is first necessary to understand behavior. In fact, training can be defined as the systematic application of psychological learning principles to facilitate the acquisition of knowledge, skills, abilities, and attitudes that result in improved performance in the desired environment. Thus, this book is about understanding both training itself and the human characteristics that underlie the process.

The basis for this book is the 9th Annual River Cities Industrial and Organizational Psychology Conference that was held on the campus of The University of Tennessee at Chattanooga in October 2013. The theme of the conference was Trends in Training. Accordingly, the focus of many of the talks was on how to train people to thrive and succeed in the environment where they spend a large portion of their lives – the workplace.

BART L. WEATHINGTON
BRIAN J. O'LEARY
CHRISTOPHER J. L. CUNNINGHAM
MICHAEL D. BIDERMAN

April 2015
Chattanooga, TN

CHAPTER ONE

CURRENT TRENDS IN RATER TRAINING: A SURVEY OF RATER TRAINING PROGRAMS IN AMERICAN ORGANIZATIONS

C. ALLEN GORMAN,

EAST TENNESSEE STATE UNIVERSITY

JOHN P. MERIAC,

UNIVERSITY OF MISSOURI-ST. LOUIS

JOSHUA L. RAY

TUSCULUM COLLEGE

AND THOMAS W. RODDY

EAST TENNESSEE STATE UNIVERSITY

Author Note:

C. Allen Gorman, Department of Management and Marketing, East Tennessee State University. John P. Meriac, Department of Psychology, University of Missouri-St. Louis. Joshua L. Ray, Department of Management, Tusculum College. Thomas W. Roddy, Department of Management and Marketing, East Tennessee State University.

Portions of this chapter were presented at the 2012 conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

We thank Caitlin Nugent, Christina Thibodeaux, Sheila List, Soniya Lonkar, Stephanie Bradley, Mamie Mason, Lindsay Pittington, and Shristi Pokhrel-Willet for their assistance with data collection.

Correspondence concerning this chapter should be addressed to C. Allen Gorman, Department of Management and Marketing, East Tennessee State University, Box 70625, Johnson City, TN, 37614. E-mail: gormanc@etsu.edu

Introduction

The evaluation of human performance in work settings has long been an interest of organizational scholars (Arvey & Murphy, 1998). Performance appraisal systems routinely require subjective ratings of performance from a worker's supervisors, peers, or subordinates. The accuracy of these ratings is important to the success of the performance appraisal system. In fact, some scholars have suggested that rating accuracy is the primary goal of performance appraisal (Werner & Bolino, 1997). Two general strategies have been advanced in the performance appraisal literature as ways of improving rating accuracy: rating scale development and rater training (Woehr & Huffcutt, 1994). Of the two, rater training has become the most widely accepted strategy (Roch, Woehr, Mishra, & Kieszczyńska, 2011). Unfortunately, however, there is no published research on the prevalence of rater training programs in organizations. The purpose of the present study is to fill this void by conducting a survey of U. S. organizations to determine if they utilize rater training and, if so, the types of rater training programs in place. In the following sections, we provide a brief overview of the rater training literature.

Rater Training

Rater training is an area that has received a substantial amount of research attention and has shown some promise for improving the accuracy of performance ratings (Woehr & Huffcutt, 1994). In general, rater training has been shown to be effective for improving the quality of performance ratings (Spool, 1978). One of the first references to rater training in the literature is credited to Bittner (1948), who noted that training provided to American army officers on the performance dimensions of the military evaluation scale improved officers' ratings of their soldiers' performance. McIntyre, Smith, and Hassett (1984) identified two major benefits of rater training: (a) to enhance raters' knowledge and skills for carrying out evaluations, and (b) to motivate raters to use the knowledge and skills learned in the training program. Perhaps an even greater benefit, Werner and Bolino (1997) found that court judges showed some preference for performance appraisal systems that included rater training programs.

Previous reviews of rater training have reported generally positive results regarding the efficacy of rater training to improve performance appraisal ratings. Spool (1978) and Smith (1986), for example, published

influential reviews of various rater training methods; however, these reviews were qualitative and did not provide cumulative statistical evidence for the effectiveness of rater training programs. To address this limitation, Woehr and Huffcutt (1994) conducted a meta-analytic review, which identified four general approaches to rater training based on their content: (a) rater error training, (b) performance dimension training, (c) frame-of-reference training, and (d) behavioral observation training. The most recent review of rater training was provided by Roch et al. (2011) in their meta-analysis of frame-of-reference training. However, as they noted, their review focused exclusively on frame-of-reference training because of a relative lack of research on other types of rater training. In the following sections, we briefly evaluate the research evidence for the four types of rater training identified by Woehr and Huffcutt (1994), and their associated limitations. Table 1 contains definitions for these rater training approaches and how they were defined to respondents in the present study.

Table 1: Rater Training Approaches and Definitions

Rater Training Approach	Survey Definition
Rater error training	Used to identify and caution raters of committing rating errors (e.g., leniency, central tendency, personal bias).
Performance dimension training	Used to provide raters with information about the performance dimensions being rated, explaining the definitions and rating scales. Raters are not given feedback concerning their actual ratings.
Frame-of-reference training	Used to provide information about performance dimensions and rating scales along with examples of positive and negative behaviors. It also includes practice sessions for the raters, followed by feedback on their ratings.
Behavioral observation training	Used to train raters to become skilled observers who can differentiate between and document poor and good behaviors of employees. It often includes training in diary keeping and note taking of employee behaviors.

Rater Error Training

Rater error training was developed as a way to combat the prevalence of psychometric errors in performance appraisal ratings (Borman, 2001). This training generally focused on halo and leniency errors (Woehr & Huffcutt, 1994). Essentially, rater error training requires raters to recognize leniency, halo, and central tendency errors and to avoid these rating patterns when making performance ratings. There is considerable evidence that rater error training reduces halo and leniency errors (see Smith, 1986, for a review of the findings of the rater error training literature). However, numerous criticisms have been leveled against this approach to rater training. For one, although rater error training results in fewer leniency and halo errors, it inadvertently lowers levels of rating accuracy (Bernardin & Pence, 1980; Borman, 1979; Landy & Farr, 1980). Others have suggested that rater error training actually produces a meaningless redistribution of ratings (Smith, 1986) and that rater errors may not in fact be errors, but rather instead could reflect true score variance (Arvey & Murphy, 1998; Hedge & Kavanagh, 1988). Arvey and Murphy (1998) summarized the current view of rater error training by suggesting that rater errors are relatively unimportant and trivial when it comes to rating accuracy. These criticisms have led to a shift in the literature towards approaches designed to improve rating accuracy (Athey & McIntyre, 1987).

Performance Dimension Training

Performance dimension training is one of several approaches that developed as a reaction to the inconsistent results of rater error training. These approaches began to emphasize the cognitive processing of information by raters as the key to the success of rater training (DeNisi, Cafferty, & Meglino, 1984; Feldman, 1981; Landy & Farr, 1983). Performance dimension training requires raters to recognize and use the appropriate dimensions of performance for which ratings will be required (Woehr & Huffcutt, 1994). This approach is based on the hypothesis that making dimension relevant judgments instead of global judgments should lead to more accurate ratings (Woehr, 1992). Training in this approach typically involves having raters review the rating scale to be used or having raters actually participate in the development of the rating scale (Woehr & Huffcutt, 1994). Research has generally found that performance dimension training is effective in improving rating accuracy (see Woehr &

Huffcutt, 1994, for a review of the research on performance dimension training).

Frame-of-Reference Training

In the same vein as performance dimension training, Bernardin and Buckley (1981) recognized the need for new rater training programs and proposed frame-of-reference training as an alternative to rater error training. As pointed out by Woehr and Huffcutt (1994), frame-of-reference training is essentially an extension of performance dimension training in that it focuses on providing raters with performance standards for each dimension to be rated, but frame-of-reference training also incorporates a practice and feedback session where raters make practice ratings and receive feedback on the quality of their ratings. The ultimate goal of frame-of-reference training is to train raters to share and use common conceptualizations of performance when making their evaluations (Athey & McIntyre, 1987; Woehr, 1994). Specifically, frame-of-reference training involves categorizing behaviors into appropriate performance dimensions as well as correctly judging the effectiveness of those behaviors (Sulsky & Day, 1992; 1994).

Many studies have demonstrated the efficacy of frame-of-reference training for improving the accuracy of performance ratings (Athey & McIntyre, 1987; Bernardin & Pence, 1980; Cardy & Keefe, 1994; Day & Sulsky, 1995; Gorman & Rentsch, 2009; Hauenstein & Foti, 1989; Hoffman, Gorman, Blair, Meriac, Overstreet, & Atchley, 2012; McIntyre, Smith, & Hassett, 1984; Noonan & Sulsky, 2001; Pulakos, 1984, 1986; Schleicher & Day, 1998; Stamoulis & Hauenstein, 1993; Sulsky & Day, 1992, 1994; Woehr, 1994). Meta-analyses of frame-of-reference training studies have found impressive effect sizes for the impact of the training on improving rating accuracy ($d = .83$ in Woehr & Huffcutt, 1994, and $d = .50$ in Roch et al., 2011). In an effort to explain why frame-of-reference training increases rating accuracy, Woehr and Feldman (1993) posited an information processing model that suggested ratings are based on both memory and contextual factors. Thus, frame-of-reference training presumably increases rating accuracy by improving these elements of performance ratings.

Although the efficacy of frame-of-reference training has repeatedly been demonstrated in the rater training literature, there are nonetheless various criticisms of this approach. One criticism focuses on the importance of memory in the rating process. Noonan and Sulsky (2001) suggested that frame-of-reference training does not instruct raters on how

to process behavior information with the goal of remembering the behavior at a later time. Indeed, memory for specific behaviors has been shown to be important when the goal is rating accuracy (Landy & Farr, 1980; Sanchez & De La Torre, 1996). A second criticism of frame-of-reference training is that the training may cause raters to see certain behaviors that were in fact never exhibited (Noonan & Sulsky, 2001; Sulsky & Day, 1992). Third, there is little attempt in frame-of-reference training research to measure the information processing that supposedly goes on as a result of the training (Arvey & Murphy, 1998; for an exception, see Gorman & Rentsch, 2009). Finally, the generalizability of frame-of-reference training research has been questioned because of its reliance on standard videotapes of performance and use of student raters in contrived laboratory settings (Arvey & Murphy, 1998; Noonan & Sulsky, 2001).

Behavioral Observation Training

As researchers began to focus on ways to improve the accuracy of ratings, new strategies emerged that emphasized the accuracy of behavioral observations. As the accuracy of behavioral observation improves, so too should the effectiveness of performance ratings (Woehr & Huffcutt, 1994). Focusing on behavior observation is important when considering that raters often must observe performance in noisy environments where competing tasks and demands deplete cognitive resources (Noonan & Sulsky, 2001). Behavioral observation training developed in response to this new emphasis on the accuracy of behavior observation. Behavioral observation training typically involves either taking notes during observations of performance or keeping a diary of observations for an extended period of time.

There is some research evidence for the effectiveness of behavioral observation training. Studies have found that behavioral observation training significantly reduces rating errors (e.g., Bernardin & Walter, 1977; Latham, Wexley, & Pursell, 1975), leads to increases in observational accuracy (e.g., Thornton & Zorich, 1980), and significantly increases rating accuracy (Hedge & Kavanagh, 1988; Noonan & Sulsky, 2001; Pulakos, 1986). However, too few primary studies have focused on the effectiveness of behavioral observation training or observational accuracy dependent measures in general (Woehr & Huffcutt, 1994). Also, as with the other rater training approaches, there are some criticisms. The main criticism of this approach is that there has been a lack of agreement regarding the conceptual definition of observation training, as well as a

lack of consensus on what exactly constitutes an observational training program (Noonan & Sulsky, 2001). A second criticism of behavioral observation training concerns its practicality, such that requiring managers to take notes and keep diaries is time-consuming and likely impractical.

Combinations of Rater Training Approaches

There has been little research investigating the impact of combined rater training strategies (Woehr & Huffcutt, 1994), and studies that have done so have been less than promising. McIntyre et al. (1984) and Pulakos (1984) noted no significant increments in rating accuracy when combining rater error and frame-of-reference training. In his review, Smith (1986) found that combinations of rater error training with other types of training failed to produce any increments in rating accuracy. Finally, both Noonan and Sulsky (2001) and Roch and O'Sullivan (2003) found that a combination of behavioral observation training and frame-of-reference training did not lead to significant increments in rating accuracy beyond frame-of-reference training alone. Overall, it appears combining rater training approaches does not add any value over and above any single training approach alone.

Summary and Conclusions of Rater Training Research

The primary conclusion that can be drawn from our review of the rater training literature is that no specific type rater training program has been shown to be *the* most effective (Hedge & Kavanagh, 1988). Given this, however, frame-of-reference training appears to be the most effective of the four general strategies for improving rating accuracy (Woehr & Huffcutt, 1994; Roch et al., 2011), although the limited generalizability of this approach may hinder its possible utility in organizations. A number of scholars have erroneously concluded that rater error training is ineffective; indeed, rater error training is effective *at reducing psychometric errors* (e.g., Latham, Wexley, & Pursell, 1975). However, as noted previously, rater error training essentially creates a meaningless redistribution of ratings and is practically useless in terms of improving rating accuracy (Borman, 1979; Smith, 1986).

Instead of arguing for or against one particular type of rater training program, we believe a more pragmatic approach for future research would be to identify the components of various training programs that appear to be the most important aspects of an effective training program. Heneman (1988), for example, suggested that rater training programs should place

equal emphasis on observation and judgment, although no single approach combines these elements in such a way. In his review, Smith (1986) observed that the successful frame-of-reference training programs all employed practice and feedback. In general, practice in making ratings and feedback about the quality of the ratings appear to be important components of any successful rater training program (Borman, 2001; Latham, 1986; Smith, 1986). As Smith (1986) concluded, the more actively involved raters become in the rating process, the greater the outcome.

Despite the promising results of rater training research in the 1980s and 1990s, there is a paucity of recent research on rater training programs compared to other topics in the field of performance appraisal. However, as noted by Roch et al. (2011) rater training is no longer an exclusively performance appraisal topic, and rater training research has expanded beyond the traditional performance appraisal context into additional rating contexts, such as the assessment center. Nonetheless, the accumulation of evidence in support of rater training for improving performance appraisal ratings suggests that rater training programs should be worthwhile interventions for improving performance ratings in organizations.

Scholars have noted, however, that despite the benefits of rater training, there remains a lack of widespread adoption of rater training programs, such as frame-of-reference training, in applied settings (Bernardin, Buckley, Tyler, & Wiese, 2001). This may be due to the potential limitations of rater training programs with regards to their application in organizations. First, rater training programs are time consuming and expensive to implement (Stamoulis & Hauenstein, 1993), and the process of developing target scores for calculating rating accuracy indices may be too complex and time-consuming for organizations (Bernardin et al., 2001; Ilgen & Favero, 1985). Second, rater training programs may also be insufficient due to low levels of user acceptance and political influence (Carroll & Schneier, 1982; Longnecker, Gioia, & Sims, 1987). Finally, beyond the cost to organizations, research has yet to show that rater training programs can generalize across jobs and members in organizations (Arvey & Murphy, 1998).

Nonetheless, research has suggested that modifications to traditional training protocols may hold promise as a means to increase organizational adoption of rater training programs. Chirico, Buckley, Wheeler, Facticeau, Bernardin, and Beu (2004), for example, provided evidence that true score feedback may not be necessary for a rater training program to be effective at improving rating accuracy. These authors compared a frame-of-reference training group that received true score feedback and another

frame-of-reference training group that received only qualitative feedback to a control group. They found that traditional frame-of-reference training and frame-of-reference training with only qualitative feedback were effective in improving rating accuracy compared to a control condition, but there were no significant differences in rating accuracy measures between the two frame-of-reference conditions. This finding suggests that rater training protocols could potentially be simplified to suit the needs of an organization while maintaining the integrity of the training principles.

The Present Study

Despite the generally positive outcomes associated with rater training, and the accumulation of evidence in support of various rater training programs, there is essentially no scholarly evidence of the prevalence of rater training programs in organizations today. There is some anecdotal evidence, however, that rater training may be increasing in popularity. In his textbook on Industrial/Organizational Psychology, for example, Levy (2010) noted that rater training has become more common in organizations such as the Tennessee Valley Authority, JP Morgan Chase, Lucent Technologies, and AT&T. In addition, a survey conducted by the Employers Resource Council (2008), found that 46% of the 73 organizations surveyed provide rater training. However, neither of these sources present any scientific evidence on the prevalence of rater training programs, nor do they provide any details on the types of training programs currently in place. Thus, we conducted an independent survey to determine the prevalence of rater training and the types of training programs currently used in organizations.

Method

Participants

Human resource executives from 101 U. S. organizations completed the survey. Titles of the executives surveyed included VP of HR, VP of Global Talent Development, Director of HR, and HR Manager. Organizations from various industries are represented in the survey, including health care facilities, medical equipment manufacturers, construction, and general merchandise. Eighty-eight percent of the 101 companies report revenues of over 1 million dollars annually, and 88% of the companies employ at least 100 employees. Most of the responding

organizations were headquartered in the Southeastern U. S. (44%), and 16% of responding companies were headquartered in the Midwestern U. S.

Procedure

Survey data were collected as part of a larger data collection effort regarding performance appraisal practices (see Gorman, Ray, Nugent, et al., 2012). We recruited HR executives to complete the survey by directly e-mailing HR departments in Fortune 500 companies, advertising the survey on popular online business forums (e.g., LinkedIn), and asking HR executives to forward the survey link to other HR executives. We specifically asked for HR executives to complete the survey because we surmised that employees in other capacities in organizations may not be aware of many of the details involved in the organization's performance appraisal system, and they may likely not understand the HR terminology associated with performance appraisal systems. Although this may not be the ideal method of data collection and does not allow reliability to be estimated, this method is consistent with other studies of HR practices such as assessment centers (e.g., Boyle, Fullerton, & Wood, 1995; Spsychalski, Quinones, Gaugler, & Pohley, 1997).

Measures

Rater training. Eight items in the survey were specific to rater training (e.g., *Does your company train managers on how to conduct performance appraisals*, and *If yes, what type of rater training is conducted [if multiple types are used, please indicate the primary method]?*). For the latter item, response options included the four rater training types identified by Woehr and Huffcutt (1994) as well as an 'Other' option (see Table 1 for the rater training definitions used in the survey).

Performance appraisal system effectiveness. To measure the perceived effectiveness of the performance appraisal system, we asked participants a single item (i.e., *Overall, how would you rate the effectiveness of your company's performance appraisal system?*). Participants responded to this item on a scale from 1 (*extremely ineffective*) to 5 (*extremely effective*).

Performance appraisal system fairness. We asked a single item to measure the perceived fairness of the performance appraisal system (i.e., *Overall, how would you rate the fairness of your company's performance*

appraisal system?). Participants responded using a scale from 1 (*extremely unfair*) to 5 (*extremely fair*).

Firm-level performance. We used a single item to measure the financial performance of the company (i.e., *Approximately how much revenue does your company make annually?*). Participants provided their response on a scale from 1 (*less than \$1 million*) to 4 (*more than \$100 million*).

Control variable. Because the implementation of rater training programs is likely to vary by the size of the company (i.e., larger companies are more likely to be able to afford to implement rater training programs than smaller ones), we used company size as a control variable in our analyses. To measure company size, we asked a single item (i.e., *How many employees does your company employ?*). Participants responded to this item using a scale from 1 (*less than 100 employees*) to 4 (*more than 500 employees*).

Table 2: Frequency Results for Rater Training and Refresher/Recalibration Training

<i>Response</i>	Train Managers	Train Managers	Non-	Refresher/Recalibration Training
Yes	77	31		50
No	24	70		19

Results

As can be seen in Table 2, 76% of the organizations surveyed indicated that they provide rater training for managers. Only 31% provide rater training for non-managers (i.e., peers, subordinates, etc.). Of the 77 organizations that provide rater training to managers, 50 (65%) also conduct refresher/recalibration training sessions. Of the 77 organizations that offer rater training for managers, the most popular type of rater training was frame-of-reference training (31 organizations; 40%), followed by performance dimension training (23 organizations; 30%). Only 13 (17%) of the 77 organizations use rater error training as the primary rater training method. Behavioral observation training was the least popular method, with only 8 (10%) of the 77 organizations that offer rater training for managers. Finally, 2 organizations (2%) indicated that they conducted other forms of training.

Table 3: Frequency Results for Who Conducts Rater Training

Training Conducted by	Frequency	Percent
External consultant	2	2.63%
Internal consultant	2	2.63%
Human resource personnel	61	80.26%
Department manager	6	7.89%
Other	5	6.58%

Note. Results are based on 76 organizations that reported they conducted rater training and provided information on who conducted it.

Table 3 shows that the overwhelming majority of rater training sessions are conducted by human resource personnel (79%). In Table 4, it can be seen that 28 (36%) of the 77 organizations that utilize managerial rater training offer the training only once per year, while 32% offer the training as needed. As one might expect, only 29% of the 77 organizations have evaluated their rater training, with the most popular criterion being trainee reactions.

Table 4: Frequency Results for Frequency of Rater Training

Frequency of Rater Training	Frequency	Percent
Less than one time a year	6	8.00%
One time per year	28	37.33%
Three times per year	13	17.33%
Three times per year	0	0.00%
Four times per year	3	4.00%
As needed	25	33.33%

Note. Results are based on 75 organizations that provided responses to this question.

Exploratory Analyses

Given the exploratory nature of our survey, we made no formal hypotheses with respect to the influence of rater training on organizational criteria. However, we conducted some exploratory analyses to examine the association between rater training and outcome variables of interest. First, controlling for company size, performance appraisal systems that utilize managerial rater training were judged to be more effective, overall, by human resource executives ($M = 3.70$, $SD = 1.51$) than performance appraisal systems that do not utilize rater training ($M = 3.37$, $SD = 1.61$),

$t(98) = 1.77, p < .05$. Although the mean fairness rating given by human resource executives was higher for performance appraisal systems that included rater training ($M = 3.58, SD = 1.04$) than those that did not ($M = 3.25, SD = 1.23$), the difference was not statistically significant, $t(98) = 0.97, ns$. In addition, controlling for company size, organizations that utilized rater training generated higher revenue ($M = 3.09, SD = 1.03$) than those that did not ($M = 2.71, SD = 1.04$), $t(98) = 3.07, p < .01$.

Although we were also interested in examining the relations between organizational outcomes and each specific type of rater training, the small sample sizes within each type precluded any meaningful analysis. However, when we combined performance dimension training, frame-of-reference training, and behavioral observation training into a higher level category we called “behavior-focused training” and regressed organizational outcomes on rater training, an interesting pattern of results emerged. Controlling for company size, as the emphasis of rater training shifted from rater error-focused to behavior-focused, the perceived effectiveness of the performance appraisal system also increased, $R(96) = .26, p < .05$. We observed the same pattern for firm-level performance such that, again controlling for company size, as the emphasis of the rater training shifted from rater error-focused to behavior-focused, the revenue generated by the company also increased, $R(96) = .41, p < .01$. No statistically significant finding was observed for perceived performance appraisal system fairness.

Supplemental Analysis

We asked the HR executives an additional question regarding the legal defensibility of their company’s performance appraisal system. Interestingly, 13% of the executives surveyed believed their current performance appraisal system was legally indefensible. We examined whether perceptions of legal defensibility depended on whether the company utilized managerial rater training. We found that performance appraisal systems were perceived as significantly more legally defensible when the system included a rater training program, $\chi^2(1, N = 101) = 4.13, p < .05$.

Discussion

In this chapter, we reviewed the rater training literature, and we provided the results of a survey designed to benchmark the current state of rater training in U. S. organizations. The results suggested that rater

training is alive and well in U. S. organizations. Seventy-six percent of the 101 organizations surveyed provide some type of rater training to managers as a part of their performance appraisal system, and 31% provide rater training to non-managers. The most popular rater training programs were frame-of-reference training (40%) and performance dimension training (30%). Most rater training programs are conducted by human resource personnel (79%) either yearly (36%) or as needed (32%), and only 29% of the surveyed organizations that offer rater training have evaluated their training programs.

The results of the present study are encouraging for several reasons. First, our results suggest that rater training is alive and well, and that, compared to the results of practitioner surveys (e.g., Employers Resource Council, 2008), rater training is on the rise in U. S. organizations. This is in contrast to the long-standing research-practice gap in performance appraisal (Banks & Murphy, 1985; Bretz, Milkovich, & Read, 1992), as well as the criticisms of the various training approaches mentioned earlier in this chapter. Rater training has been a topic of study in the performance appraisal literature for many years and continues to receive a fair amount of research attention. And, although our data does not shed any light on trends in rater training over the years, our results did confirm that rater training is currently a popular intervention for improving ratings in U. S. organizations. We were also encouraged that 31% of the organizations in our survey provide rater training to non-managers. Given the popularity of multisource performance appraisals (i.e., 360-degree ratings) in organizations, it is important to recognize that anyone who will be evaluating the job performance of another employee receive training on the proper methods to do so, especially if the ratings will be used for administrative purposes. However, because multisource performance appraisals were originally designed for developmental purposes (rather than administrative), improving the accuracy of these ratings may not be the primary concern to practitioners (Hoffman et al., 2012).

Second, frame-of-reference training and performance dimension training are the “go to” rater training approaches in organizations today. This is encouraging because these training methods are behavior-focused, rather than error-focused, and behavior-focused methods lead to an emphasis in organizations on the accuracy of ratings rather than the supposed errors that a rater could potentially make (Bernardin & Buckley, 1981). Indeed, we found that rater error training is waning in popularity, perhaps deservedly so given the inconsistent results of rater error training in the literature. In fact, our results suggest that performance appraisal systems that employ rater error training are perceived as not as effective as

behavior-focused training types, and we provided preliminary evidence that behavior-focused methods are associated with higher company revenue, although the evidence is correlational and causation should not be inferred. We were, however, a bit surprised by the lack of adoption of behavioral observation training programs, although the emphasis on diary-keeping and note-taking in this approach may not lend itself to widespread practical application.

Third, organizations should be encouraged that rater training has been widely adopted across many different organizations and industries. With the accumulation of support for rater training in the literature, combined with the evidence from our survey, we see no reason that rater training should not be a mainstay in every performance appraisal system. As Werner and Bolino (1997) noted, judges show a preference for performance appraisal systems that have a rater training component. Moreover, our results showed that human resource executives perceive performance appraisal systems without rater training as less effective and less legally defensible (although there were no significant differences for perceived fairness), and companies with no rater training programs tend to do worse financially. One could argue that, given the legal implications and the body of evidence supporting rater training, providing rater training to managers should be an ethical imperative for organizations. That is, given the negative consequences of not providing rater training to managers and the benefits associated with providing rater training, human resource executives potentially face a moral obligation to provide rater training to all managers in their organizations. Human resource executives should continue to encourage and support rater training programs in their organizations, particularly those that emphasize the evaluation of job behaviors.

Areas for Improvement

Despite the positive findings regarding the prevalence of rater training in organizations, our survey identified several areas that remain fertile ground for improvement. First, we found that only 22 of the 77 organizations that offer rater training have evaluated the training. This is not surprising, however, given that the majority of training programs, in general, are never evaluated (Goldstein & Ford, 2002). The results from our survey suggest that what little evaluation has been done on rater training programs has been primarily using reaction criteria. Although reactions convey some information regarding the utility of the training, they tell you very little regarding the amount of learning that has taken

place as a result of the training or whether the knowledge and skills learned will apply to the job (Baldwin & Ford, 1988; Blume, Ford, Baldwin, & Huang, 2010). Only limited research exists that has investigated the transfer of rater training to the practice of performance appraisal on the job (Macan, Mehner, Havill, Roberts, Heft, & Meriac, 2011). In the study conducted by Macan et al., managers who were trained as assessors in an assessment center using frame of reference training provided significantly more behavioral statements in performance appraisals compared with untrained managers. Future research should continue to evaluate the extent that rater training transfers to the practice of performance appraisal on the job, the ways that different types of rater training transfer, and the specific ways training impacts performance ratings.

Second, we found that the majority of rater training sessions are provided either once per year or as needed. Although this is not surprising given that most organizations conduct their performance appraisals only once per year (Gorman et al., 2012), human resource personnel should be aware of the potential for the decay of knowledge and skills learned during training. Meta-analytic evidence suggests that knowledge and skills imparted through training decay over time without practice and rehearsal (Arthur, Bennett, Stanush, & McNelly, 1998). Thus, with only one training session and only one opportunity per year to display their knowledge and skills, managers are likely not retaining all of the knowledge and skills learned during rater training. We were, however, encouraged that 50 of the 77 organizations that offer rater training also offer refresher/recalibration training. We are aware of no research that has examined the impact of refresher/recalibration training on the quality of performance ratings.

Limitations and Future Research Directions

As with any study, there are potential limitations associated with the present study. First, the survey relied on single source data, with human resource executives providing all of the information regarding their organization's performance appraisal system and the outcome measures. Although the observed relationships could potentially be inflated due to common method bias, the rater training items were only a small part of the overall survey regarding the organization's performance appraisal system. Thus, it is unlikely that the executives were thinking specifically of rater training when providing their perceptions of their performance appraisal systems, although we did not examine this possibility in the present study.

Still, future research could benefit from using multiple sources in responses (i.e., system users as managers/raters and subordinates/ratees). Additionally, more information on user reactions could be informative. Specifically, managers' user reactions may be improved by experiencing rater training, such that they feel more confident in the perceived usefulness or experience less discomfort when making performance ratings. Ratees could also report more positive user reactions to the extent that training improves the quality of ratings by increasing the specificity of information in reported behaviors and accuracy.

Second, the linkages between rater training and financial outcomes were exploratory and based on cross-sectional data, and no causal inferences can be made. It could also be that the more financially well-off institutions are the only companies that are able to afford the expense associated with rater training programs. And there could also plausibly be a third variable driving this relationship. Moreover, these results were obtained only after controlling for company size; thus, our findings here cannot be interpreted in isolation. Changes over time could be studied in experimental, longitudinal or pre-post designs to more directly study changes as a result of implementing rater training. Nonetheless, we were intrigued by these findings, and we encourage other researchers to continue to explore the mechanisms behind the influence of rater training on firm-level outcomes using longitudinal research designs.

Third, an argument could be made that the rater training approaches used in the organizations we surveyed were not completely identical to how these approaches have been described and tested in the literature. For example, a frame-of-reference training program employed in one organization might not contain all of the same elements as a frame-of-reference training program in another organization. Further, it is possible that organizations may incorporate some elements of multiple training approaches but only recognize it as one specific form. We do not know for sure either way, but we did ask the executives to indicate their organization's primary training approach using the definitions culled from the literature. Thus, it is possible that the organizations may actually use a combination of different training approaches. Future research in this area should directly assess the specific elements of training (e.g., practice, feedback), the duration of training, and other characteristics to better understand exactly how training is implemented. It is also possible that some organizations implement forms of training that do not fall cleanly into one of the four categories that were included in this study. Although only two organizations checked an "other" category, other respondents may have checked the closest available category instead. Further, it is

likely that some organizations incorporate “system use” training for an online system that is not actually rater training as studied here. However, as mentioned previously in this chapter, we believe the success of rater training is due to the underlying elements of the training, such as practice and feedback, rather than the label assigned to the particular training approach. Nonetheless, we encourage researchers to further explore this area to determine whether organizations are employing single rater training approaches or combinations of various approaches, and if the latter, what are the most successful combinations. We implore researchers to continue to survey organizations regarding their use of rater training programs, and we urge practitioners to partner with researchers in designing and evaluating rater training programs in their organizations. Rater training represents one of the most effective approaches to improving the quality of performance ratings, but relatively little information is available regarding the extent to which practitioners have incorporated findings from the research literature.

In the present study, we focused on companies in the U.S., based on a sample generalizable to other U.S. companies. While our sample size limited our ability to do so, future research could also explore whether the use of rater training varies by industry or company strategy, as well as how rater training fits with organizational goals and missions. Further, the extent to which rater training is implemented in other countries remains an open question. Finally, we encourage the examination of the transfer of rater training in organizations in addition to studying the long term impact of rater training on organizational results and outcomes.

Conclusion

To our knowledge, the survey presented in this chapter is the first independent survey of rater training programs in U. S. organizations. We found that rater training is a popular intervention for improving ratings in organizations, and its popularity seems to be increasing. Although preliminary, the results of the present study also suggest that behavior-focused rater training programs increase the perceived effectiveness of the organization’s performance appraisal system, and they are associated with increased company performance. Our exploratory findings, combined with the research reviewed in this chapter, suggests that rater training is a worthwhile investment for organizations and, perhaps, it should become a standard component of all performance appraisal systems.

References

- Arthur, W., Bennett, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*, 57-101. doi: 10.1207/s15327043hup1101_3
- Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49*, 141-168. doi: 10.1146/annurev.psych.49.1.141
- Athey, T. R., & McIntyre, R. M. (1987). Effect of rater training on rater accuracy: Level of processing theory and social facilitation theory perspectives. *Journal of Applied Psychology, 72*, 239-244. doi: 10.1037/0021-9010.72.4.567
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*, 63-105. doi: 10.1111/j.1744-6570.1988.tb00632.x
- Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. *Personnel Psychology, 38*, 335-345. doi: 10.1111/j.1744-6570.1985.tb00551.x
- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review, 6*, 205-212. doi: 10.5465/AMR.1981.4287782
- Bernardin, H. J., Buckley, M. R., Tyler, C. L., & Wiese, D. S. (2000). A reconsideration of strategies for rater training. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (pp. 221-274). Stamford, CT: JAI Press Inc.
- Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology, 65*, 60-66. doi: 1981-06507-001
- Bernardin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology, 62*, 64-69. doi: 1977-20093-001
- Bittner, R. H. (1948). Developing an industrial merit rating procedure. *Personnel Psychology, 1*, 403-432. doi: 10.1111/j.1744-6570.1948.tb01319.x
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management, 36*, 1065-1105. doi: 10.1177/0149206309352880
- Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology, 64*(4), 410. doi: 10.1037/0021-9010.64.4.410

- . (2001). Job behavior, performance, and effectiveness. In Dunnette, M. D. & Hough, L. M. (Eds.), *Handbook of industrial and organizational psychology*, Vol. 2 (2nd ed.) (pp. 271-326). Palo Alto, CA, US: Consulting Psychologists Press.
- Boyle, S., Fullerton, J., & Wood, R. (1995). Do assessment/development centres use optimum evaluation procedures? A survey of practices in UK organizations. *International Journal of Selection and Assessment*, 3, 132-140. doi: 10.1111/j.1468-2389.1995.tb00017.x
- Bretz, R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management*, 18, 321-352. doi: 10.1177/014920639201800206
- Cardy, R. L., & Keefe, T. J. (1994). Observational purpose and evaluative articulation in frame-of-reference training: The effects of alternative processing models on rating accuracy. *Organizational Behavior and Human Decision Processes*, 57, 338-357. doi: 10.1006/obhd.1994.1019
- Carroll, S. J., & Schneier, C. E. (1982). *Performance and review systems: The identification, measurement, and development of performance in organizations*. Dallas, TX: Scott, Foresman, and Company.
- Chirico, K. E., Buckley, M. R., Wheeler, A. R., Fecteau, J. D., Bernardin, H. J., & Beu, D. S. (2004). A note on the need for true scores in frame-of-reference (FOR) training research. *Journal of Managerial Issues*, 16, 382-395.
- Day, D. V., & Sulsky, L. M. (1995). Effects of frame-of-reference training and information configuration on memory organization and rating accuracy. *Journal of Applied Psychology*, 80, 001-009. doi: 10.1037/0021-9010.80.1.158
- DeNisi, A. S., Cafferty, T., & Meglino, B. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Decision Processes*, 33, 360-396. doi: 10.1016/0030-5073(84)90029-1
- Employers Resource Council (2008, August). Smart business workplace practices survey. Retrieved from <http://www.ernet.org/research/studies/08ERCsBNWorkplacePracticesSurveyreport.pdf>.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127-148. doi: 10.1037/0021-9010.66.2.127
- Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations*. Belmont, CA: Wadsworth.

- Gorman, C. A., Ray, J., Nugent, C., Thibodeaux, C. N., List, S., Lonkar, S., Bradley, S., Mason, M., Pittington, L., & Pokhrel-Willet, S. (2012, April). *A preliminary survey of performance management practices in the US*. Poster presented at the 2012 conference of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Gorman, C. A., & Rentsch, J. R. (2009). Evaluating frame-of-reference rater training effectiveness using performance schema accuracy. *Journal of Applied Psychology, 94*, 1336-1344. doi: 10.1037/a0016476
- Hauenstein, N. M. A., & Foti, R. J. (1989). From laboratory to practice: Neglected issues in implementing frame-of-reference rater training. *Personnel Psychology, 42*, 359-379. doi: 10.1111/j.1744-6570.1989.tb00663.x
- Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology, 73*, 68-73. doi: 10.1037/0021-9010.73.1.68
- Heneman, R. L. (1988). Traits, behaviors, and rater training: Some unexpected results. *Human Performance, 1*, 85-98. doi: 10.1207/s15327043hup0102_1
- Hoffman, B. J., Gorman, C. A., Blair, C. A., Meriac, J. P., Overstreet, B. L., & Atchley, E. K. (2012). Evidence for the effectiveness of an alternative multisource performance rating methodology. *Personnel Psychology, 65*, 531-563. doi: 10.1111/j.1744-6570.2012.01252.x
- Ilgen, D. R., & Favero, J. L. (1985). Limits in generalization from psychological research to performance appraisal processes. *Academy of Management Review, 10*, 311-321. doi: 10.2307/257972
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*, 72-107. doi: 10.1037/0033-2909.87.1.72
- Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance: Methods, theory, and applications*. New York: Academic Press.
- Latham, G. P. (1986). Job performance and appraisal. *International Review of Industrial and Organizational Psychology, 1*, 117-53. doi: 10.4102/sajip.v12i1.452
- Latham, G. P., Wexley, K. N., & Pursell, F. D. (1975). Training managers to minimize rating errors in the observation of behavior. *Journal of Applied Psychology, 60*, 550-555. doi: 1976-01674-001
- Levy, P.E. (2010). *Industrial/Organizational psychology* (3rd ed). New York, NY: Worth Publishers.

- Longnecker, C. O., Gioia, D. A., & Sims, H. P. (1987). Behind the mask: The politics of employee appraisal. *Academy of Management Executive, 1*, 183-193. doi: 10.5465/ame.1987.4275731
- Macan, T., Mehner, K., Havill, L., Roberts, L., Heft, L., & Meriac, J. P. (2011). Two for the price of one: Assessment center training to focus on behaviors can transfer to performance appraisals. *Human Performance, 24*, 443-457. doi: 10.1080/08959285.2011.614664
- McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and purpose of rating. *Journal of Applied Psychology, 69*, 147-156. doi: 10.1037/0021-9010.69.1.147
- Noonan, L. E., & Sulsky, L. M. (2001). Impact of frame-of-reference and behavioral observation training on alternative training effectiveness criteria in a Canadian military sample. *Human Performance, 14*, 3-26. doi: 10.1207/S15327043HUP1401_02
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology, 69*, 581-588. doi: 10.1037/0021-9010.69.4.581
- . (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational Behavior and Human Decision Processes, 38*, 78-91. doi: 10.1016/0749-5978(86)90027-0
- Roch, S. G., & O'Sullivan, B. J. (2003). Frame of reference rater training issues: recall, time and behavior observation training. *International Journal of Training and Development, 7*, 93-107. doi: 10.1111/1468-2419.00174
- Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczyńska (2011). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology, 85*, 370-395. doi: 10.1111/j.2044-8325.2011.02045.x
- Sanchez, J. I., & De La Torre, P. (1996). A second look at the relationship between rating and behavioral accuracy in performance appraisal. *Journal of Applied Psychology, 81*, 3-10. doi: 10.1037/0021-9010.81.1.3
- Schleicher, D. J., & Day, D. V. (1998). A cognitive evaluation of frame-of-reference training: Content and process issues. *Organizational Behavior and Human Decision Processes, 38*, 78-91. doi: 10.1006/obhd.1998.2751
- Smith, D. E. (1986). Training programs for performance appraisal: A review. *Academy of Management Review, 11*, 22-40. doi: