The Trainer Matters: Cross-Classified Models of Trainee Reactions

DAVID R. GLERUM
The Ohio State University

DANA L. JOSEPH
University of Central Florida

AARON F. MCKENNY
Indiana University

BARBARA A. FRITZSCHE
University of Central Florida

Author Note: This article is based on the thesis completed by Glerum (2011). A version of this article was presented at the 31st annual conference of the Society for Industrial and Organizational Psychology and at the 2016 conference of the International Personnel Assessment Council. This research was supported by an award from the International Personnel Assessment Council.
This article is dedicated to the memory of William A. Wooten. We thank Martha E. Hennen, Daniel A. Newman, Deniz S. Ones, and Kevin Yee for their valuable feedback on an earlier version of this manuscript. We would also like to thank Ali Ferguson for assistance in manuscript preparation and copyediting.

Correspondence concerning this article should be addressed to David Glerum, Fisher Leadership Initiative, The Ohio State University, 300 Fisher Hall, 2100 Neil Avenue, Columbus OH 43210; email: glerum.3@osu.edu

Abstract

Despite the common belief that "training is only as effective as the trainer providing it" (Osborn, 2018, para. 1), training theory tends to underemphasize the trainer and instead focuses on training content and design as sources of training effectiveness. In this article, we examine whether the role of the trainer should be more central to training theory. We address this issue using a dataset of trainee reactions from more than 10,000 employees enrolled in professional development courses. We suggest that trainee reactions are more likely to be influenced by the trainer than by the content. Thus, trainee reactions should reflect more between-trainer variance than between-content variance. Across two studies in online and face-to-face contexts, cross-classified random-effects models provide general support for our hypotheses, with one notable exception: the trainer matters less for trainee reactions in online courses. Our findings suggest the trainer matters more than previously thought, and thus, training theory should incorporate the role of the trainer in training effectiveness. Based on our findings, we suggest that training researchers should (a) model the trainer as a source of variation in training evaluation metrics, (b) examine the effect of the trainer at multiple levels of analysis, and (c) explicitly model the role of the trainer in training theory and design.

Keywords: training evaluation, training effectiveness, trainee reactions, cross-classified models, computer-aided text analysis

The Trainer Matters: Cross-Classified Models of Trainee Reactions

Training plays a critical role in equipping employees to do their jobs and is a potential source of competitive advantage (Bell, Tannenbaum, Ford, Noe, & Kraiger, 2017; Noe, Clarke, & Klein, 2014; Tzabbar, Tzafrir, & Baruch, 2017). Given the significance of training, many organizations evaluate training effectiveness, defined as "the extent to which trainees (and their organization) benefit [from training] as intended" (Brown & Sitzmann, 2011, p. 486). Kirkpatrick's (1956, 1996) training evaluation framework suggests that training effectiveness should be evaluated on four criteria: trainee reactions to the training, knowledge or skill acquisition, transfer of trained skills to the job, and results (e.g., increased sales).

Organizations enhance training effectiveness by identifying characteristics of the training that predict these training evaluation criteria. Traditional (Baldwin & Ford, 1988; Cannon-Bowers, Salas, Tannenbaum, & Mathieu, 1995; Goldstein, 1993) and multilevel (Mathieu & Tesluk, 2010; Sitzmann & Weinhardt, 2018) theories of training effectiveness suggest that training effectiveness is determined by the trainee, the work environment, and the training design. Notably underemphasized among this list of determinants is the trainer (Brown & Sitzmann, 2011). Formally defined, a trainer is an individual who "helps people gain new skills, knowledge, or behaviors; acquire proficiency and awareness of products, processes, or methods; and achieve a defined or perhaps higher performance standard" (Association for Talent Development, 2016, p. 1). When models of training effectiveness *do include* the trainer, the role ascribed to the trainer is often narrow. For instance, Tannenbaum, Cannon-Bowers, Salas, and Mathieu (1993) include the trainer in their model of training effectiveness, but they focus primarily on the trainer's role in influencing trainee self-efficacy and not on the other ways in which trainers influences trainees. Surprisingly, the trainer does not play a more prominent role

in the training literature given the common perception that "training is only as effective as the trainer providing it" (Bahn, 1973; Kaman, 1985; Osborn, 2018, para. 1).

In contrast, training design has been given substantial consideration (Baldwin & Ford, 1988; Cannon-Bowers et al., 1995; Kozlowski, Chao, & Jensen, 2010). For instance, course content is present in many training theories as a key determinant of training effectiveness and is defined as the topical material covered in a training course, including the knowledge and skills to be learned or improved (e.g., Baldwin & Ford, 1988; Burke & Hutchins, 2007; Kraiger, 2002). Course content that is job-relevant and timely is more likely to be used by trainees than that which is irrelevant and dated (Burke & Hutchins, 2008). In one classic example, a training director was challenged with designing a training course for millworker supervisors on the mill's union contract (Thayer & McGehee, 1977). Eventually, the organization abandoned the course after discovering the managers already knew the material and would be unmotivated if the course were held.

In this study, we evaluate and compare the roles of the trainer and training course content in influencing trainee reactions, which are defined as trainees' evaluations of their training experiences (Sitzmann, Brown, Casper, Ely, & Zimmerman, 2008). Because trainee reactions are the most collected form of effectiveness data (Patel, 2010), the determinants of trainee reactions have implications for key training decisions. When trainee reactions are poor, organizational leaders must answer the question of what should be changed—the trainer, the content, or both. Answering this question may inform training research by clarifying the roles of the trainer and the course content in influencing trainee reactions. Answering this question may also inform training practice by providing suggestions for how to utilize and interpret trainee reactions effectively.

In this article, we make four contributions to the training literature. First, we examine the extent to which trainee reactions are driven by the trainer and by the course content using cross-classified models of trainee reactions from a national professional development company. Second, we compare the importance of the trainer to the importance of the course content to determine which matters more in forming reactions. Third, because the number of online courses is increasing (Derouin, Fritzsche, & Salas, 2005), we evaluate whether our findings differ across face-to-face courses, where the trainer is physically present during training, and online courses, where the trainer virtually facilitates training. Finally, we use the findings to address how theories of training effectiveness should incorporate the trainer.

Sources of Trainee Reactions Variance

Although trainee reactions have modest utility in predicting whether trainees will successfully apply what they learn to their jobs (Colquitt, LePine, & Noe, 2000; Ruona, Leimbach, Holton, & Bates, 2002; Sitzmann, Brown, Casper, Ely, & Zimmerman, 2008), they facilitate trainee motivation and self-efficacy following training (Sitzmann et al., 2008). Moreover, organizations can use trainee reactions data to decide what courses to invest in, how to design courses, whom to select as trainers, and what feedback to give trainers (Brown & Gerhardt, 2002; Kraiger, 2002; Sitzmann et al., 2008). Given the popularity and practicality of trainee reactions, we chose to examine trainee reactions as an indicator of training effectiveness.

A vast literature on reactions to educational courses has demonstrated that the course content and the trainer serve as primary sources of trainee reactions (Andersen, 1979; Gorham, 1988). Indeed, Moore's (1989, 1997) work on student reactions suggests that learners interact primarily with these two factors. Similarly, most training experiences involve a trainer interacting with trainees to deliver knowledge and develop skills (Goldstein & Ford, 2002).

Therefore, we borrow from the educational psychology literature to suggest that the trainer and the course content are the most fundamental sources of variance in trainee reactions.

The Trainer

The educational psychology literature suggests that the trainer is *the* most important factor influencing trainee achievement (Meyer, 2001; Sanders, 2000). Indeed, scholars have described the influence of instructors on learning outcomes as "strong and large" (Odden, Borman, & Fermanich, 2004, p. 7). Research has demonstrated that instructor characteristics, such as personality and influence tactics, are strongly related to course evaluations (Darling-Hammond, 2000; Pounder, 2007; Seidel & Shavelson, 2007). Furthermore, recent meta-analyses have suggested that the instructor influences achievement by fostering student engagement (Roorda, Jak, Zee, Oort, & Koomen, 2017; Roorda, Koomen, Spilt, & Oort, 2011), classroom management (Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016), and direct effects on learning (Kyriakides, Christoforou, & Charalambous, 2013).

Although educational research has emphasized the instructor, organizational research on training has underemphasized the trainer. Yet, some work indicates that trainer characteristics influence trainee reactions (e.g., Sanders & Rivers, 1996; Sitzmann & Weinhardt, 2018). For example, Sitzmann et al. (2008) have found that trainer immediacy, defined as the extent to which the trainer uses communication behaviors that reduce the distance between him or her and the learner, is a stronger predictor of trainee reactions than a number of trainee characteristics (e.g., mastery goal orientation, agreeableness, etc.). Additionally, Towler and Dipboye (2001) have found that trainees react more positively to expressive trainers than inexpressive trainers. These studies provide valuable preliminary evidence that the trainer influences trainee reactions.

The Training Content

Training content comprises the facts, concepts, procedures, rules, and principles unique to a training course (Wulfeck, Ellis, Richards, Wood, & Merrill, 1978). Effective training ensures that training content is relevant to or mirrors the tasks, activities, and considerations encountered in the focal trainee's profession (Baldwin & Ford, 1988; Gagné, 1962). As such, training content plays a central role in traditional models of training effectiveness (Baldwin & Ford, 1988; Cannon-Bowers et al., 1995). Training content is expected to influence trainee reactions because aligning training content to on-the-job content directly affects trainees' utility judgments. Furthermore, training content can affect trainees' receptivity to learning, attentional focus, information processing, and metacognition (Gully & Chen, 2010; Mathieu, Tannenbaum, & Salas, 1992). Thus, research suggests that content is a determinant of trainee reactions.

Comparing the Trainer and the Content as Sources of Trainee Reactions Variance

Cross-classification of trainee reactions. If trainee reactions are poor, what should be changed: the trainer or the content? Which matters more? Unfortunately, this type of research question has not been examined by training scholars, perhaps because it is difficult to obtain the data required to answer it. Answering this question requires a dataset of trainee reactions from multiple trainers who have taught multiple training courses. In addition, the content needs to be standardized across trainers to ensure the effects of the trainer and the content are as independent as possible. The influence of the trainer and the training content can then be compared using cross-classified models.

Cross-classified models are used when observations are nested within two or more categories that do not subsume each other. For example, trainees may enroll in courses that vary in both content (e.g., diversity training, interpersonal skills training, technical training) and

trainer (e.g., Trainer A, Trainer B, Trainer C). As such, trainees' reactions may be *cross-classified* by the course content and the trainer categories. Cross-classified models allow us to compare variance in trainee reactions due to the trainer (i.e., between-trainer variance) to those due to the training content (i.e., between-content variance). As a result, we can test (a) whether the trainer accounts for variance in trainee reactions, suggesting training theory should incorporate the role of the trainer; (b) whether the course content accounts for variance in trainee reactions, confirming that training theory should continue to include course content; and (c) which accounts for more variance: the trainer or the content. These three tests offer valuable insights for practitioners who use trainee reactions to inform training design.

We argue that overall trainee reactions are likely to reflect more between-trainer variance than between-content variance because the entire training experience is funneled through the trainer. Specifically, we expect that trainees focus on their experiences with the trainer as the primary source of event-based cues when providing their reactions to a course. As eye-tracking studies have suggested, trainers play an important role in facilitating cognitive engagement and directing trainees' attention toward course content (Moreno, Reislein, & Ozogul, 2010; Ouwehand, van Gog, & Paas, 2015). In these studies, trainees tended to direct their attention toward the trainer during the lecture rather than the content-relevant features of the course, such as the textbook or slides. Thus, overall trainee reactions are expected to reflect more between-trainer variance than between-content variance, because the trainer serves as the primary conduit for the training content. As such, we propose the following:

> *Hypothesis 1: Trainee reactions reflect more between-trainer variance than between-content variance.*

Although we hypothesize that overall trainee reactions reflect more between-trainer variance than between-content variance, we believe that the dimensions of trainee reactions may

reflect different sources of variance. Trainee reactions are generally thought to comprise two key dimensions: *satisfaction* and *utility* (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Brown, 2005; Tracey, Hinkin, Tannenbaum, & Mathieu, 2001). Satisfaction is defined as a reaction reflecting whether trainees liked and enjoyed the training (Alliger et al., 1997), and utility is defined as a reaction reflecting the perceived usefulness, practicality, and relevance of the training (Alliger et al., 1997). In the following sections, we discuss hypothesized sources of variance in satisfaction and utility reactions.

Satisfaction reactions reflect between-trainer variance. In terms of determining whether an individual liked or enjoyed a training course, the affect infusion model suggests that individuals are more likely to rely on their feelings rather than on concrete objective information (Forgas & George, 2001). Similarly, the affect-as-information model suggests that when individuals make satisfaction evaluations, they are essentially asking themselves, "How do I *feel* about the training?" (Schwarz, 1990). As such, individuals will primarily draw from the source (e.g., the trainer or the course content) most closely tied to their affect when making these judgments.

Although trainees may focus on the trainer when making satisfaction judgments, they may focus less on the course content because it lacks emotional substance. In most training courses, course content, such as how to use a software package or drive a forklift, is affectively neutral. According to the affect infusion model, individuals rely on course content less than affectively laden information when forming judgments (Forgas & George, 2001; Rocklage & Fazio, 2016, 2018). In contrast, the affect infusion model also implies that the trainer is more likely than the content to influence satisfaction reactions because experiences with a trainer are naturally affective and are thus more accessible in memory than neutral information. For

example, social-cognitive approaches to student reactions in educational psychology suggest that the instructor is a key figure who influences students' emotional experiences, such as enjoyment (Frenzel, Becker-Kurz, Pekrun, Goetz, & Lüdtke, 2018; Keller, Hoy, Goetz, & Frenzel, 2016; Pekrun, 2006). Therefore, we argue that satisfaction reactions are more likely to reflect between-trainer than between-content variance because satisfaction reactions rely on affect (Howardson & Behrend, 2016), and this affective information is more likely to be conveyed through trainees' recollections of experiences with the trainer than with the course content.

Satisfaction reactions may also reflect more between-trainer variance than between-content variance because they evoke relational attributions associated with the trainer-trainee relationship. Recent theoretical advancements in attribution theory define relational attributions as explanations for an event based on the relationship an individual (e.g., a trainee) has with another person, such as receiving a poor evaluation because "the trainer does not like me" (Eberly, Holley, Johnson, & Mitchell, 2011, 2017). As noted by Eberly et al. (2011, 2017), individuals are likely to make relational attributions when they depend on another individual (see also, Kelley & Thibaut, 1978; Rusbult & Van Lange, 2008). Given that trainees depend greatly on the trainer for enjoyable training experiences, they may relationally attribute the lack of an enjoyable experience to the trainer. Indeed, prior research in educational psychology has provided evidence for dyadic effects between students and teachers on student reactions, demonstrating that students form idiosyncratic impressions of how well they get along with a teacher or benefit from their teaching style (Göllner, Wagner, Eccles, & Trautwein, 2018).

Overall, satisfaction reactions are more likely to reflect between-trainer than between-content variance because (a) satisfaction reactions are judgments that call on more affective sources of information, such as the trainer as opposed to the course content, and (b) the training

experience is heavily dependent on the trainer, which can activate relational attributions. Therefore, we propose the following:

> *Hypothesis 2: Trainee satisfaction reactions reflect more between-trainer variance than between-content variance.*

Utility reactions reflect between-content variance. Although we argue that the trainer plays a bigger role than the course content in trainee satisfaction and overall reactions, we suggest that training content plays a bigger role than the trainer in trainee utility reactions. Utility reactions represent evaluations of the relevance, practical value, and usefulness of a training course to trainees' jobs or personal/professional development (Warr & Bunce, 1995; Warr et al., 1999). Trainees' evaluations of utility are likely affected by whether the training content covers knowledge that is necessary for trainees' jobs. If the training content addresses a real need, trainees generally evaluate the training as more useful (Webster & Martocchio, 1995). For example, trainees are likely to perceive a training course that covers how to use a new software program required to do their jobs as useful. However, a training course that covers how to report unethical behavior, such as whistleblower training, may not be perceived as useful to trainees who do not need this information to do their jobs. Noe (1986) suggests that another way training content affects perceived utility is through its instrumentality for improving promotion/raise opportunities and increasing status and prestige. As such, trainees may also see content that helps them grow professionally as useful or relevant.

Moreover, we argue that utility reactions reflect between-content variance because trainees often perceive that course content is outside the trainer's control (Steiner et al., 1991). Research generally supports this assertion, suggesting that trainees view training content as separate from controllable learning-process elements related to trainer performance (e.g., training styles, organization of the material, effective time management, effective use of learning

activities, etc.; Holladay & Quiñones, 2008; Morris, 1984; Towler & Dipboye, 2001). For example, Towler and Dipboye (2001) have found that trainees actively process training content even when the trainer delivers the content in a disorganized way. In sum, trainees' utility reactions are more likely to reflect between-content variance because (a) utility perceptions reflect how important the course content is for trainees' jobs or professional development and (b) trainees may perceive the utility of the training course as outside the trainer's control. Therefore, we propose the following:

> *Hypothesis 3: Trainee utility reactions reflect more between-content variance than between-trainer variance.*

Overview of Studies

We obtained anonymous trainee reactions data from a professional development company to test our hypotheses. The company offered 559 sections of 31 graduate-level training courses for schoolteachers in both online and face-to-face training formats. The trainees needed to pass the courses to fulfill licensure requirements and to apply for salary increases. The courses covered topics like preventing bullying, differentiating instruction, educating special needs students, service learning, and multicultural education. The trainees completed reactions surveys at the end of each course.

The trainee reactions data were collected by the professional development company between 2009 and 2010 and were nested within both content and trainers.[1] Of the 13,506 trainee reactions solicited, 11,892 trainee reaction surveys were returned (88.05% response rate). We excluded 1,613 responses because they were from exclusively online or face-to-face courses.

[1] To obtain more information regarding the attributes of the 31 course offerings included in this study, five graduate students were asked to rate the workload of each course by examining the course syllabi. A single workload item was rated as deficient (1), sufficient (2), or substantial (3). The courses were masked by title and content to circumvent any subject- or content-related biases the raters might have had. The raters were provided training on how to rate the syllabi on perceived workload and were directed to focus on the breadth of workload, difficulty level, and level of time investment outside of class. The graduate students rated these syllabi with acceptable to good levels of agreement (ICC(3, 5) = .62; Cicchetti, 1994). Workload levels were moderate and varied modestly across courses (M = 2.12, SD = 0.40), suggesting there was some similarity in workload across courses.

Retaining courses offered in both training formats enabled us to evaluate whether format influences the hypothesized relationships. In addition to the analysis of quantitative Likert-type trainee reactions, we used content analysis to examine the sources of variance in qualitative open-ended comments (e.g., Harman, Ellington, Surface, & Thompson, 2015). As such, one aim of both studies is to compare whether our results vary as a function of quantitative vs. qualitative indicators of trainee reactions.

The trainees consisted of kindergarten through high school teachers, speech instructors, college-level instructors, adult education instructors, and English language learner instructors who taught subjects including science, gifted courses, physical education, music, and art. The majority (98.53%) of the sample declared a teaching occupation, whereas a small portion declared other occupations, such as “administrator” and “librarian.” The trainees did not need to take these courses with this company: the teachers could choose to take courses with any credit-granting institution accepted by their district. The courses were offered in partnership with universities and were held to a similar standard as graduate-level education. The trainees were aware of their training course grades (A, B, C, or fail) only after providing their trainee reactions.

There were 58 trainers who facilitated these courses, including 15 who taught both online and face-to-face courses. Most of the trainers were female (64.79%). All the trainers held a master’s degree or higher, had been pre-approved by academic partner universities as eligible for adjunct faculty positions, and were certified K–12 classroom teachers. Notably, the training course content was standardized across the delivery formats and the trainers, with nearly identical course objectives, materials, timelines, and outlines within each course (with slight modifications in the online format). The trainees were required to complete a final project as a part of their grade. The trainers could not adapt specific course content for their training sessions.

On average, each trainer taught approximately nine courses over the two-year data-collection period.

Study 1

Sample and Method

Only trainees who enrolled in face-to-face courses were included in Study 1. This sub-sample consisted of 7,097 trainees enrolled in 280 sections across several states in the United States. Forty-five different trainers taught face-to-face sessions across 31 courses. On average, each course contained 25.35 trainees (*SD* = 12.73), and each trainer taught an average of 157.71 trainees (*SD* = 151.51).

Likert-type trainee reactions. The satisfaction reactions scale consisted of five items asking trainees to rate whether the trainer exhibited (a) preparation/organization, (b) knowledge of the subject, (c) creation of a positive learning atmosphere, (d) equal concern for theory and practice, and (e) effective classroom-management skills. The utility reactions scale consisted of five items asking the trainees to rate (a) the course in comparison with previous training courses they had taken, (b) the usefulness of the textbook, (c) the usefulness of the workbook, (d) the usefulness of the experiential activities, and (e) how well the course applied to their classroom. We averaged all 10 items from the satisfaction and utility scales to construct an overall evaluation scale. The items were rated on a 5-point Likert-type rating scale ranging from 1 (*low*) to 5 (*high*) and demonstrated excellent internal consistency (α = .94 for overall evaluation; α = .96 for satisfaction; α = .89 for utility).

These items are similar to those included in typical trainee reactions scales. For example, the items and scale format are similar to Towler and Dipboye's (2001) trainee reactions scale, which asks trainees to rate the trainer on a series of adjectives and behaviors (e.g., "competent,"

"friendly," etc.) and the course on a variety of characteristics (e.g., "usefulness," "interest," etc.). Furthermore, we analyzed 127 published primary studies we obtained from Sitzmann et al.'s (2008) meta-analysis of trainee reactions and found that 85.12% included a satisfaction scale, 56.20% included a utility scale, and 79.34% involved a brief survey ($M_{items} = 9$) similar to that of the current study. Unfortunately, no one trainee reactions scale is used ubiquitously in the training evaluation literature, and most scales are adapted on a study-by-study basis (Brown & Sitzmann, 2011). As such, although the current scale was developed for use by the professional development organization, we believe that it is similar to others' scales and that the items appropriately assess the content domain of trainee reactions.

For the Likert-type trainee reactions data, we tested the hypotheses utilizing cross-classified random-effects models (CCREMs; Meyers, 2012) with maximum likelihood estimation (MLE; Harville, 1977). We used the Satterthwaite (1946) approximation to specify the degrees of freedom for each intercept parameter estimate in the unconditional model. To test each hypothesis, we calculated an intraclass correlation coefficient (ICC; Raudenbush & Bryk, 2002) for each of the cross-classified factors using an extension of the ICC formula for CCREMs (Beretvas, 2011),

$$\rho = \frac{\sigma^2_{u_{0j_1}}}{\sigma^2_e + \sigma^2_{u_{0j_1}} + \sigma^2_{u_{0j_2}} + \sigma^2_{u_{0j_1} * u_{0j_2}}} \quad (1)$$

which is a function of each variance component for the *j*th cross-classified random effect, their interaction (which is often treated as error; Webb, Shavelson, & Haertel, 2006), and the residual variance component. We also expressed the ICCs as percentages for each cross-classified effect and compared them with one another.

Trainee comments. In addition to the trainee reactions scales discussed above, the survey also included "Comments" fields after both the satisfaction and utility item blocks to enable the trainees to comment on their perceived satisfaction and utility ($N_{Satisfaction}$ = 3,192; $N_{Utility}$ = 1,901). The trainees provided a total of 5,093 comments. Using dictionary-based computer-aided text analysis (CATA; McKenny, Aguinis, Short, & Anglin, 2018), we analyzed these qualitative comments to facilitate additional comparison of between-trainer vs. between-content variance. Dictionary-based CATA is a form of content analysis in which a computer identifies and counts words and short phrases (hereafter, "words") in a text, and the frequency of these words reflects psychological aspects of the text's author (Duriau, Reger, & Pfarrer, 2007). With regard to our study, we examined the trainees' satisfaction and utility reactions using CATA measures (see "Supplemental Material" for the dictionary, development, and validation of these measures).

To test our hypotheses using the trainee comments data, we used a Bayesian cross-classified zero-inflated Poisson (CCZIP) model to estimate the appropriate variance components for the qualitative trainee reactions (Chen, Brown, & Stokes, 2016; Liu & Cela, 2008). We chose this method because the data were in count form, there was an over-abundance of zeros (Blevins, Tsang, & Spain, 2015), and to improve the estimation of the variance components (LoPilato, Carter, & Wang, 2015). We used predictive quasi-likelihood (PQL; Hox, 2010) estimation, a form of MLE, to estimate the variance components. Given that comment length varied across the trainees and qualitative items, it was necessary to include an offset variable. This variable was the total word count per comment and was log-transformed to be on the same scale as the count outcome variable (Hox, 2010). We performed the analyses with at least 100,000 burn-in iterations (Kruschke, Aguinis, & Joo, 2012) that were thinned to result in a posterior sample of

10,000. Following Gelman (2006), we sampled hyperpriors for all the variance components from the data using uniform distributions ranging from 0 to 5. Following best practices for Bayesian estimation of variance components, we chose the half-Cauchy distribution centered on the standard deviation for the variance component priors (Gelman, 2006). We did not use informative or empirical priors (LoPilato et al., 2015), which are typically drawn from prior research, as this is the first study to examine cross-classified models of qualitative trainee reactions.

Results

Confirmatory factor analysis. The means, standard deviations, and correlations among the trainee reactions items are presented in Table 1. First, we conducted a confirmatory factor analysis (CFA) with MLE to test whether the hypothesized two-factor (satisfaction and utility) model fit the data better than alternative models. To test which model was most appropriate, we evaluated multiple goodness-of-fit indices, examined overlap in the root-mean-square error of approximation (RMSEA) 90% confidence intervals (CIs), and conducted change in Comparative Fit Index tests (ΔCFI; West, Taylor, & Wu, 2012). The factor loadings and fit indices are presented in Table 2. The hypothesized model with two correlated latent factors fit the data relatively well. The utility and satisfaction latent factors were correlated (φ = .81), suggesting considerable overlap among the two factors. The hypothesized model fit better than a single-factor model (ΔCFI = .099; $RMSEA_{2\text{-factor}}$ = .108, 90% CI = [.105, .112]; $RMSEA_{1\text{-factor}}$ = .177, 90% CI = [.174, .180]). A higher-order model subsuming both the utility and satisfaction factors fit nearly identically, but this model included a Heywood case, suggesting misspecification. In sum, the factor analysis results support our use of a two-factor model representing satisfaction and utility. As such, we averaged the items to form satisfaction and utility scales, respectively.

However, given that aggregated scales are often used in practice, we included aggregated scales as well.

Cross-classified random-effects models: Likert-type trainee reactions. The results for the unconditional CCREMs are reported in Table 3. Hypothesis 1 suggested that overall trainee reactions reflect greater between-trainer variance than between-content variance. Hypothesis 1 was supported as the variance in 90.00% of the trainee reactions items was primarily attributable to the trainer. Hypothesis 2 suggested that satisfaction reactions reflect between-trainer variance more than between-content variance. All satisfaction items reflected more between-trainer variance than between-content variance, supporting Hypothesis 2. The third hypothesis suggested that utility trainee reactions reflect between-content variance more than between-trainer variance. All but one utility item (textbook usefulness) reflected more between-trainer variance than between-content variance. This result suggests that Hypothesis 3 was only supported by one of the five items. We observed the same pattern of results at the scale level as the satisfaction scale and the utility scale reflected more between-trainer variance than between-content variance. Overall, Hypothesis 3 was not supported.

Cross-classified zero-inflated Poisson models: Open-ended trainee comments. The unconditional Bayesian CCZIP models of the open-ended comment data are presented in Table 4. Table 4 is broken down by comment item and CATA trainee reactions dimension. Table 4 includes the modal variance components as the mode tends to be a more accurate representation of central tendency for variance components than the mean (LoPilato et al., 2015). The table also includes the 95% credibility interval, which establishes the 2.5% and 97.5% boundaries of the posterior distribution for the variance component parameter, providing a fixed range within which the estimated modal posterior variance component lies. Finally, the ICCs are presented to

compare the trainer and the course content as sources of variance in trainee comments. To determine the larger source of variation, we set a greater than 5.00% difference in ICC as a conservative cutoff for variance component disparities because significance tests are traditionally a component of frequentist, and not Bayesian, approaches (Gelman & Shalizi, 2013).

Regarding the CATA satisfaction reactions, the trainer was the larger source of variance for the open-ended satisfaction item. This result provides support for Hypothesis 2 (satisfaction reactions reflect more between-trainer variance than between-content variance). Regarding Hypothesis 3 (utility reactions reflect more between-content variance than between-trainer variance), the CATA utility reactions reflected between-content variance, as predicted, for the open-ended utility item. We also examined the extent to which the CATA satisfaction reactions reflected between-trainer and between-content variance when applied to the open-ended utility item. In this instance, the trainer was at least an equivalent source of variance to the content for the open-ended utility item, and the trainer × content interaction reflected the largest amount of variance. However, the CATA utility reactions unexpectedly reflected between-trainer variance when applied to the open-ended satisfaction item. Furthermore, a large amount of the variance in the CATA utility reactions was reflected in the trainer × content interaction.

Discussion

We found support for our hypothesis that overall trainee reactions reflect more between-trainer variance than between-content variance (Hypothesis 1). Similarly, our results largely supported our hypothesis that satisfaction reactions reflect more between-trainer variance than between-content variance (Hypothesis 2). The only exception was for the most conservative test of this hypothesis, which indicated that the CATA satisfaction reactions reflected equal levels of trainer- and between-content variance for the open-ended *utility* item. In contrast, our expectation

that utility reactions would reflect more between-content variance was generally not supported. Overall, the trainer mattered more than we expected: of the 14 tests in which we compared between-trainer and between-content variance, 11 returned more between-trainer variance, one returned equal portions of variance, and two returned more between-content variance. The lack of between-content variance was particularly surprising for some of the Likert-type utility items that were more directly aligned with the course content, such as items asking respondents about the workbook and course activities.

Even though these results provide a compelling initial test of our hypotheses, there are several limitations to this study. First, it is not clear whether the findings generalize to other types of training courses—for example, online courses. Furthermore, the Likert-type satisfaction items explicitly referenced the trainer, which may present a confound if the items reflect between-trainer variance because trainees were prompted to focus on the trainer. Notably, this confound was not an issue for the Likert-type utility items, which tended to reflect more between-trainer variance despite referring to training content. Similarly, because the open-ended satisfaction item was presented immediately after the Likert-type satisfaction items, the trainees may have been primed to reflect on the trainer more than they otherwise would have if the comment field did not appear immediately after items referencing the trainer. We address these limitations in Study 2.

Study 2

Overview

Study 2 builds upon the results of Study 1 in three ways. First, and most importantly, Study 2 addresses the confound of the satisfaction scale in Study 1, in which the trainer was included as an item referent, by using measures of trainee reactions that are unaffected by this

confound. For instance, we examined the variance sources of an "overall rating" that did not reference the trainer as an alternative test of Hypothesis 1. Overall ratings are a global assessment of trainee reactions used often in research and practice (Alliger & Horowitz, 1989). Moreover, we examined the variance sources of additional open-ended comment items that were not presented in association with any of the Likert-type scales and did not mention the trainer or course content. These alternative tests of our hypotheses were neither confounded with Likert-type language nor the trainer referent confound in Study 1.

Second, Study 2 builds on Study 1 by using a broader range of item content that reflects several types of trainee reactions measures used in practice. Thus, we can examine the extent to which our findings generalize to other measures. For instance, many of the scales included training characteristics important to many online courses (e.g., reactions toward e-mail interaction, the discussion forum, etc.).

Third, Study 2 examines the generalizability of the findings from Study 1's face-to-face context to online trainer-aided courses with the same training content. We tested our hypotheses in this context to determine if the dimensions of trainee reactions reflect the same sources of variance or if they are altered by contextual differences, such as transactional distance. This contribution is important as training courses are increasingly being delivered online (Derouin et al., 2005; Noe et al., 2014). Prior meta-analytic research has suggested that the differences between trainee reactions in face-to-face and online training are very small (Sitzmann, Kraiger, Stewart, & Wisher, 2006). However, research on transactional distance has suggested that online learning fosters psychological distance and communication gaps that may limit the interaction between trainers and trainees (Chen, 2001; Moore, 1989, 1997), which could influence trainee reactions. In addition, research on both verbal and nonverbal communication has suggested that

trainers engage in less adaptive instruction online, so that they are less responsive to trainees than they would be if they interacted in person (e.g., Offir, Lev, Lev, Barth, & Shteinbok, 2004). Furthermore, trainer immediacy, which is defined as trainees' perceptions of physical and psychological closeness fostered by nonverbal behavior (Mehrabian, 1981), may be reduced through an online medium because it cannot fully transmit tone, body language, and other communication nuances experienced in person. Given the transactional distance and reduced immediacy between the trainer and the trainee in online courses, we believe online trainee reactions may reflect less between-trainer variance than face-to-face courses.

Sample and Method

Trainees enrolled in 279 online training courses completed trainee reactions at the end of their course (N = 3,182). Twenty-eight trainers taught online sessions across 31 courses. On average, each online course contained approximately 11.41 trainees (SD = 7.50). To encourage class participation, the trainers served as "moderators" on class discussion boards, leading discussion and posing questions, and the trainees were required to thoughtfully participate in these activities as a part of their grade. The trainers had a virtual "open-door" policy and engaged in one-on-one discussions electronically with students in a similar fashion as face-to-face trainers. The trainees were required to post their typed final project (e.g., as a Word document) to their course's discussion board, which served as the basis for discussion for the other trainees. Essentially, interaction with the trainer in the online context duplicated what would be expected in the face-to-face context except for the lack of physical synchronous interaction.

Likert-type trainee reactions. The trainees completed the reactions survey at the end of their final course module. The satisfaction reactions scale consisted of seven items asking trainees to rate whether the trainer exhibited (a) preparation/organization, (b) knowledge of the

subject, (c) creation of a positive learning atmosphere, (d) equal concern for theory and practice, (e) email interaction, (f) accessibility/responsiveness, and (g) warmth/rapport. We averaged these items to construct an overall satisfaction reactions scale. The utility reactions scale consisted of five items asking trainees to rate (a) the course in comparison with previous training courses they had taken, (b) the usefulness of the textbook, (c) the usefulness of the website materials, (d) the usefulness of the discussion forum, and (e) how well the course applied to their classroom. We averaged these items to construct an overall utility reactions scale. Finally, the survey included an item capturing the trainees' "overall rating of this course" to assess global reactions. We averaged all 12 items to construct an overall evaluation scale. All items were rated on a 5-point Likert-type rating scale ranging from 1 (*low*) to 5 (*high*) and demonstrated good to excellent internal consistency (α = .92 for overall evaluation; α = .95 for satisfaction; α = .83 for utility).

Trainee comments. The online trainees were presented with seven open-ended items that did not refer to a specific target, such as the trainer. A "Comments" box appeared directly after the "overall rating of this course" item (N = 502). A similar comment box was also positioned immediately after the final satisfaction item (N = 753). In addition to these two general "Comment" field items, the trainees completed five specific open-ended items: (a) "What are the strengths of this course and online delivery?" (N = 2,624), (b) "What are the weaknesses of this course and online delivery?" (N = 2,354), (c) "What would you change about this course?" (N = 2,296), (d) "Any other comments about the online course experience?" (N = 1,463), and (e) "Please describe the results you experienced personally and professionally in this course" (N = 2,265). In total, we analyzed 12,257 comments using the CATA dictionaries developed in Study 1.

Results

Confirmatory factor analysis. The means, standard deviations, and correlations among the trainee reactions items are presented in Table 1, and the CFA results are presented in Table 2. The hypothesized model with two correlated latent factors fit the Likert-type data relatively well. The utility and satisfaction latent factors were correlated (φ = .70), suggesting considerable overlap among the two factors. The hypothesized model also fit the data better than a single-factor model (ΔCFI = .149; $RMSEA_{2\text{-factor}}$ = .088, 90% CI = [.084, .092]; $RMSEA_{1\text{-factor}}$ = .183, 90% CI = [.179, .186]). A higher-order model subsuming both utility and satisfaction factors fit nearly identically, but this model included a Heywood case, suggesting misspecification. In sum, the results support a two-factor model representing separate satisfaction and utility scores.

Cross-classified random-effects models: Likert-type trainee reactions. The results for the unconditional CCREMs are reported in Table 5. Hypothesis 1 suggested that aggregate overall reactions primarily reflect between-trainer variance. We found that most trainee reactions reflected both between-trainer and between-content variance to a similar degree, although most items reflected more trainer variance. These estimates were very similar in magnitude, and much of the variance in trainee reactions remained unexplained. To test this hypothesis with a different indicator of overall reactions, we examined whether a global single-item rating primarily reflects between-trainer variance. This hypothesis was not supported as 2.60% of the variance was between-content, whereas 1.10% of the variance was between-trainer, estimates that were very similar in magnitude. Although there were inconsistencies in whether trainee reactions reflected between-trainer or between-content variance across scale level, item level, and single-item indicators, the estimates were similar in magnitude and mostly unexplained. Regarding Hypothesis 2, which proposed that satisfaction reactions primarily reflect between-trainer

variance, six of the seven items reflected more between-trainer variance than between-content variance despite small differences in magnitude. Hypothesis 3 suggested that utility trainee reactions reflect between-content variance more than between-trainer variance. At the item level, although between-content variance was the larger source of variance for all but one item, the average differences between the sources of variance were very small (3.44% on average). Across all analyses, trainee reactions appeared to reflect both trainer and content variance to a roughly equivalent degree (a difference in magnitude of less than 5.00%).

Cross-classified zero-inflated Poisson models: Open-ended trainee comments. The unconditional CCZIP models for the online courses are presented in Table 6, and a summary of the CATA results can be found in Table 7. We could not completely test Hypothesis 1 using the CATA trainee reactions because we did not have a CATA measure of global reactions. However, when we concurrently examined the CATA satisfaction and utility reactions toward the five general open-ended items that did not refer to satisfaction or utility, there were 10 unconfounded opportunities to assess whether the CATA reactions reflected more between trainer or between content variance. The results showed more between-trainer variance for 50.00% of these tests, equivalent amounts of between-trainer and between-content variance for 50.00% of these tests, and more between-content variance for 0.00% of these tests.

Regarding Hypothesis 2, which posited that satisfaction reactions reflect more between-trainer variance than between-content variance, we found that the CATA satisfaction reactions reflected more between-trainer variance than between-content variance. The trainer was the larger source of variance for the CATA satisfaction reactions in three of the seven open-ended items and was at least equivalent to the course content as a source of variance in three of the seven open-ended items. In only one instance did a CATA satisfaction reaction reflect more

between-content variance: the open-ended "results experienced" item, which most closely aligns with utility. It should be noted that in all the cases in which the between-trainer variance was not larger than the other source, there was a large trainer × content interaction component. This result suggests that CATA satisfaction reactions reflect the interaction between the trainer and the course for open-ended items.

We did not find support for Hypothesis 3, which proposed that utility measures of trainee reactions reflect more between-content variance than between-trainer variance. None of the CATA utility reactions in the open-ended items reflected more between-content variance than between-trainer variance. On the contrary, four out of the seven items reflected more between-trainer variance than between-content variance, with the remaining three items reflecting an equivalent degree of trainer and content variance. The "results experienced" item, which most closely represents an open-ended elicitation of utility reactions, reflected more between-trainer variance than between-content variance. As with the CATA satisfaction reactions, a substantial trainer × content interaction variance component was generally observed for the CATA utility reactions.

Discussion

Across all analyses in Study 2, the differences between the trainer and the content were very small. Overall, our quantitative and qualitative results suggest that the trainer tends to capture as much or more variance than the course content with one exception: the global single-item reflected more between-content variance. Thus, we conclude general support for Hypothesis 1 in the online sample, noting that there were differences in support depending upon the reactions measure. Global items that many organizations use to assess trainee reactions may

operate differently than multi-item scales or CATA measures, as they do reflect more between-content variance than between-trainer variance.

In partial support of Hypothesis 2, both the quantitative Likert-type satisfaction items and the CATA satisfaction reactions tended to reflect more between-trainer variance than between-content content variance or at least an equivalent degree of both sources of variance. Of the seven satisfaction Likert-type items and the seven open-ended items analyzed with the CATA satisfaction measure, only two reflected more between-content variance than between-trainer variance. Furthermore, in the most direct test of Hypothesis 2 in the qualitative analyses, the CATA satisfaction reactions reflected more between-trainer variance than between-content variance for the open-ended satisfaction item and were unconfounded by direct reference to the trainer. These results suggest that trainee satisfaction is influenced more by the trainer than by the course content although, trainee reactions often reflect both trainer and content variance in online courses.

Finally, our analyses did not provide consistent support for Hypothesis 3. The Likert-type utility reactions primarily reflected between-content variance, whereas the CATA measure of utility failed to reflect more between-content variance than between-trainer variance, including the item that appeared to most directly assess utility. This result suggests that unless targeted questions about utility are asked, such as the Likert-type questions used in Study 2, trainee utility reactions may largely reflect the trainer rather than the course content. This finding has implications for measuring trainee reactions, which we discuss in our general discussion below.

General Discussion

Across two studies and a combined sample of more than 10,000 Likert-type trainee reactions and more than 17,000 comment-based trainee reactions, our results suggest that the

trainer matters more than the training content for trainee reactions. Our findings are important to training theory, in which the role of the trainer is often underemphasized. Together, our findings call attention to the need for training theory to consider the role of the trainer in training evaluation. These findings are also important for organizational decision-makers. They suggest (a) that if trainee reactions are poor, the trainer is more likely the source of the problem than the course content and (b) that even when trainee reactions specifically note that the course content is a problem, changing the course content might not improve trainee reactions. Below, we discuss our results in more detail, noting how our findings vary across face-to-face and online training.

As Studies 1 and 2 involved multiple opportunities to assess our hypotheses, we summarize the results in Table 8, showing percentages of how often each item reflected between-trainer variance, between-content variance, or a roughly equal amount of both in Studies 1 and 2. We also summarize the support for the hypotheses in Table 9. Although many of the cells in Tables 8 and 9 suggest that support for the hypotheses was moderated by measure type, we note that the overall pattern of results led to three conclusions: (a) across all analyses, measures, and studies, between-trainer variance was the largest source of variance, with 50.00% of analyses indicating more between-trainer variance than between-content variance (only 7.50% of the analyses indicated more between-content variance, suggesting that the trainer matters more than the content for trainee reactions); (b) the role of the trainer is not as strong in online training as it is in face-to-face training given that Hypotheses 1 and 2 received stronger support in Study 1 than in Study 2; and (c) unexpectedly, utility judgments do not tend to reflect the course content more than the trainer, as Hypothesis 3 proposed.

Implications for Training

Updating training effectiveness theory. Goldstein and Ford (2002) note, "Training is a people-to-people activity" (p. 4). Our results support this notion, demonstrating the importance of the trainer for trainee reactions and offering several implications for training effectiveness theory. First, most traditional training effectiveness frameworks underemphasize the role of the trainer (e.g., Baldwin & Ford, 1988). Our study suggests that these frameworks may benefit from emphasizing the trainer as a more central figure in the training process. Second, the substantial between-trainer variance we found implies that trainer characteristics are important as they may influence trainee reactions. However, we do not yet understand *which* trainer characteristics (e.g., intelligence, communication ability, personality) are most influential for trainee satisfaction and utility, suggesting an opportunity for further cross-level research. For example, Grossman and Salas (2011) suggest that trainer persuasiveness is critical for convincing trainees of the usefulness and relevance of training, which can facilitate mastery goal orientations and improve transfer of training. Third, our finding that the trainer influenced trainee reactions provides an impetus to examine whether trainer characteristics may be an underexamined influence on other indicators of training effectiveness (Kirkpatrick, 1956, 1996; Kraiger, Ford, & Salas, 1993).

Future research should tease apart the degree to which trainer characteristics (e.g., personality; Kim, Dar-Nimrod, & MacCann, 2018) vs. trainer behaviors (e.g., pedagogical style, expressiveness; Towler & Dipboye, 2001) comprise between-trainer variability. Such work would have implications for both training and selection; namely, should organizations devote more effort into selecting trainers with particularly desirable personality traits, or should organizations devote time toward ensuring trainers learn and use proper instructional styles? Furthermore, future research could draw on multilevel training evaluation theory (Mathieu &

Tesluk, 2010; Sitzmann & Weinhardt, 2018) to more closely examine episodic trainer-trainee interactions and how these interactions lead to the formation of satisfaction and utility judgments. For example, prior research on cheerleading instructors demonstrates that the situation plays a role in influencing trainer performance and how well they deliver the training (Beal, Trougakos, Weiss, & Green, 2006; Trougakos, Beal, Green, & Weiss, 2008), but understanding how these influences affect trainee reactions over time would be invaluable to our understanding of trainee reactions. Studies such as these may help unpack how trainee reactions form and further inform relational attributions theory, the affect infusion model, and the affect-as-information model.

Transactional distance and immediacy in online courses. Although between-trainer variance was the largest source of variance in face-to-face courses, this was not the case in online courses. Instead, trainee reactions tended to reflect an equal amount of between-trainer and between-content variance, although much of the variance in online reactions was unexplained. Nevertheless, ratings of "warmth/rapport" and "accessibility/responsiveness" largely reflected between-trainer variance in online courses, despite how other characteristics appeared to be suppressed by the online delivery format. These results align with findings from the communication immediacy literature, which suggest that online trainers' individual differences influence how well they engage online classes (Thomas & Thorpe, 2019). Certain online instructional features (e.g., group chat, message boards, e-mail) may also help communicate information in this context (Thomas & Thorpe, 2019). Thus, perhaps training design features and engaging trainers can circumvent the challenges imposed by transactional distance and lack of immediacy.

The reduced importance of the trainer in the online context also supports transactional distance explanations of why online and face-to-face training differ in effectiveness. Transactional distance theory suggests that the psychological distance in online courses limits effective communication and the extent to which individual trainer differences can manifest (Moore, 1989, 1997). Our findings have implications for training effectiveness theory in general, suggesting that delivery format changes the training process by reducing trainer immediacy and increasing transactional distance. Training effectiveness theory should recognize the potential limit of the trainer's reach in online training courses. Reactions in this context likely reflect other sources of variance. As such, training theory should investigate delivery format as a potential moderator of the trainer-training effectiveness relationship.

Finally, an untested assumption is that trainees are aware of what the trainer is responsible for in a training course. The extent of responsibility might be relatively clear in the face-to-face context, whereas the trainer's responsibility may be more ambiguous and difficult for trainees to determine in online courses due to increased transactional distance and a lack of communication immediacy. This lack of awareness of what the trainer is responsible for in the training courses may have resulted in the similar between-trainer and between-content variance observed in Study 2. As such, future research should more explicitly test how trainees attribute the trainer's responsibility for different course components in face-to-face and online courses. Developing trainee reactions measures that more precisely assess perceptions of responsibility for course components would be particularly helpful in such an effort.

Edutainment in training. Unexpectedly, the trainees' use of satisfaction language in their responses to course/utility comments reflected a large portion of trainer × course content interaction variance in both studies, suggesting that the way the trainer delivers the content

matters. On the one hand, some trainers may be better at facilitating certain courses than other trainers, reflecting skill-based differences in trainer performance for certain content. On the other hand, the trainer × course content interaction could also highlight trainers' use of edutainment practices (Beato, 2015; Billsberry, 2014; Okan, 2003). Edutainment involves presenting "interesting and entertaining information that is irrelevant or only marginally related to the intended theme of the course," including "seductive details," such as humorous jokes, video clips, and stories (Harp & Mayer, 1997; Sitzmann & Johnson, 2014, p. 1). Sitzmann and Johnson (2014) have found that trainers' use of edutainment is a double-edged sword: although this tactic reduces trainees' negative affect during the training course, it also causes them to direct less attention toward the course content, hindering their learning performance. In our study, multiple trainees within a course section commented that their trainer's use of "music and humor really add to the class," that the trainer "makes the class fun," and that the trainer's anecdotes were "entertaining." However, one trainee commented on the irrelevance of this trainer's materials: "I am very disappointed in the instructor. Although the instructor was very nice and pleasant to talk to, there was not enough learning. The instructor talked too much about their experiences and their school district." While entertaining, these edutainment practices may draw attention away from the course content.

Implications for the Measurement of Trainee Reactions

Sources of variance in trainee reactions. Future research should examine additional sources of variance in trainee reactions. Indeed, a large amount of the variance in our analyses across both studies remained unexplained, suggesting that there may be other unmeasured sources of variance. Future research would benefit from modeling the trainee as a source of variance (Gully & Chen, 2010; Mathieu et al., 1992), especially in online courses as trainee

preferences for online courses vary substantially. A comparison of the trainee, the trainer, and the content as sources of variance in trainee reactions would be quite informative. For example, if the trainee is the source of most of the variance, trainee reactions may reflect individual differences in trainee personality or affectivity (e.g., Göllner et al., 2018). In this study, trainee reactions were anonymous, so data on individual differences were not collected, limiting the extent to which we could model the trainee as a random effect.

In addition, given that online courses may be offered more frequently than face-to-face courses, trainers may receive feedback on their course performance more frequently than in face-to-face courses and may adjust their pedagogy or style accordingly. Adding time as an additional source of variance (e.g., modeling courses or trainers over time) may further clarify the extent to which trainee reactions reflect improvements in training courses or trainers over time (e.g., Wagner, Göllner, Werth, Voss, Schmitz, & Trautwein, 2016) although we did not find time to be a prominent source of variation in our data.[2]

Furthermore, the trainee × time interaction may be a particularly interesting source of variance to examine how much variance may be accounted for by growth in trainees' reactions over each consecutive course they take. Along with our earlier recommendation to examine how trainee reactions are formed within training episodes, researchers should also examine additional sources of variance throughout the duration of training courses, using experience sampling methods (Larson & Csikszentmihalyi, 1983). For example, recent research in educational

[2] We added time as another form of cross-classification (along with the three additional interaction effects) and found that in all cases in Study 2, trainee reactions did not significantly reflect time-relevant variation. However, in the face-to-face data from Study 1, the trainer × time interaction was significant for the overall aggregated reactions [ICC(1) = .107], utility [ICC(1) = .107], and satisfaction scales [ICC(1) = .089]. Although not relevant to our question of comparing between-trainer with between-content sources of variance, trainee reactions did reflect (at least in face-to-face courses) changes in trainer effectiveness over time, with approximately 9.00–11.00% of the variance reflecting this source. Incorporating this source of variation did not change the substantive conclusions of Study 1, and the trainer was still the primary source of variation when compared to the content.

psychology suggests that rater severity drift (i.e., escalating severity in ratings of instruction over time) can have an impact on reactions over time (Casabianca, Lockwood, & McCaffrey, 2015; Congdon & McQueen, 2000; Harik, Clauser, Grabovsky, Nungester, Swanson, & Nanakumar, 2009; Wilson & Case, 2000).

Likert-type vs. CATA measures of trainee reactions. Our analysis of qualitative trainee reactions data helped provide complementary, nuanced trainee perceptions that may not be captured in Likert-type scales, corroborating Harman et al.'s (2015) work on open-ended comments. Furthermore, the trainee reactions were self-generated. This open-ended format allowed the trainees to respond to the items using their own language, adding to the quality of trainee reactions measurement. Although our use of CATA measures constitutes a different approach than that used by Harman et al. (2015), we were able to provide evidence supporting the generalization of between-trainer variance to open-ended comments. Future research should examine sources of variance in other measures of trainee reactions, such as comment characteristics (e.g., tone, scope, and purpose) and commenting behavior.

Across both studies, very similar patterns of results emerged for the Likert-type and CATA measures of trainee reactions (see Table 8). Between-trainer variance was the largest source for both measures. Notably, it appears that between-content variance tended to be much larger for certain items, such as textbook and workbook usefulness, which may be because the textbook and workbook are much more closely linked to the training course content than other course aspects. As such, practitioners developing trainee reactions measures should be aware of the extent to which each item is linked to the respective source of variance.

Like Kraiger (2002), we argue that trainee reactions measures should align with the purpose of training evaluation, such as revising future training or developing trainers, and closely

focus on the essential aspects of training design. At the same time, we recognize that measures that are idiosyncratic to one context limit the comparability of findings across contexts. To facilitate comparability across contexts, we encourage researchers and practitioners to develop measures of trainee reactions that incorporate elements from both approaches. We encourage the development and validation of psychometrically sound scales for measuring trainee reactions that can be applied across contexts. Developing scales such as these would enable studies to examine the generalizability of our findings and extend them further to address more nuanced questions. At the same time, we advocate complementing scales that can be applied across contexts with scales catered to the purpose and context of the training.

We offer the following recommendations to accomplish this aim. First, ensure that the trainee reactions measure aligns with the purpose of training evaluation, such as to provide feedback to trainers or decide which aspects of the content to revise. Second, exercise caution in making the purposes of formative training evaluation transparent because trainees may strategically respond to reactions measures to further their own goals, such as rating difficult content as less useful to reduce their workload. Finally, place emphasis on the content validity of the trainee reactions measure: the content of the measure should align with the training content, which in turn should align with what the trainees encounter on their jobs or in their professions. This final point is especially important: if trainees believe that practice exercises are the most useful part of the training course, but the reactions measure does not target practice exercises, the measure may be deficient and not accurately represent trainee utility reactions.

Limitations

One limitation of this study is that the course content and the trainer may not capture the universe of sources of variation in trainee reactions. However, researchers should strive for

parsimony in CCREMs by focusing only on the critical nesting factors in question in any given application (Meyers, 2012). In both educational and organizational psychology, course content and trainer variables comprise the two primary points of interaction in learning events (Andersen, 1979; Gorham, 1988; Kraiger, 2002; Moore, 1989). We believe that the criticality of these two factors is also reflected in our data: for example, one trainee noted, "I would've liked more time for practical classroom application and sometimes more defined directions (not sure if it's more the instructor or the course itself)." This excerpt illustrates that the trainee was focusing on either the trainer or the course content in their attribution.

Another limitation can be found in the trainee reactions instrument itself: the partner organization developed the instrument, and explicit information on development procedures was not available. However, the trainee reactions scales utilized in this study evidenced construct validity and were very similar to those used in prior studies in length and content. Given the similar structure and content of our trainee reactions measures to those used in prior work, we believe that our results can be replicated using different measures. We also addressed this limitation by using a variety of trainee reactions measure types, such as the CATA measures.

We also note that the face-to-face and online courses were not precisely the same across delivery formats. For example, the trainees interacted asynchronously via e-mail and discussion boards in online courses. As such, although the organization made efforts to standardize the experience across delivery modes as much as possible, some of the training practices may not have been entirely reproduced online. Furthermore, given the asynchronous nature of the online courses, they required more time to complete than the face-to-face courses. Future research should attempt to replicate this work in contexts with a higher level of virtuality, such as in synchronous, virtual, instructor-led training. Finally, the trainees self-selected into courses of

their choosing, so future studies should seek samples in which trainees are randomly assigned to online or face-to-face courses.

Conclusion

Although training effectiveness theory typically does not emphasize the role of the trainer, our results suggest the trainer influences trainee reactions. Our findings indicate that the trainer influences both satisfaction and utility reactions—even to a greater degree than the training content. One way trainers might affect trainees is through their delivery of the course content, a notion that aligns with previous research on edutainment. Furthermore, the trainer's influence appears to be modified by whether the course is conducted online or face to face. In online courses, reactions appear to be less affected by the trainer, as communication immediacy and transactional distance theories suggest. As such, training theory should account for the limited reach the trainer may have in the online context due to increased psychological distance. Our results also echo the idea that "training is a people-to-people activity" (Goldstein and Ford, 2002, p. 4) and suggest that research and practice should continue to uncover individual trainer differences as well as elements of the trainer-trainee relationship that influence training effectiveness. In general, our results suggest that training theory should be updated to incorporate the role of the trainer, and we call on future work to uncover the trainer characteristics underlying this substantial between-trainer variance.

References

Alliger, G. M., & Horowitz, H. M. (1989). IBM takes the guessing out of testing. *Training & Development Journal*, *43*, 69–73.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, *50*(2), 341–358. https://doi.org/10.1111/j.1744-6570.1997.tb00911.x

Andersen, J. F. (1979). Teacher immediacy as a predictor of teaching effectiveness. In D. Nimmo (Ed.), *Communication Yearbook* (Vol. 3, pp. 543–559). New Brunswick, N.J.: Transaction Books.

Association for Talent Development. (2016, January). Basic Training for Trainers. *TD at Work*, *33*(1613).

Bahn, C. (1973). The counter training problem. *Personnel Journal*, *52*(12), 1068–1072.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: a review and directions for future research. *Personnel Psychology*, *41*(1), 63–105. https://doi.org/10/cq3btj

Beal, D. J., Trougakos, J. P., Weiss, H. M., & Green, S. G. (2006). Episodic processes in emotional labor: Perceptions of affective delivery and regulation strategies. *Journal of Applied Psychology*, *91*(5), 1053-1065. https://doi.org/10.1037/0021-9010.91.5.1053

Beato, G. (2015, March 19). That's Edutainment. *The New York Times*, p. F1.

Bell, B. S., Tannenbaum, S. I., Ford, J. K., Noe, R. A., & Kraiger, K. (2017). 100 years of training and development research: What we know and where we should go. *Journal of Applied Psychology*, *102*(3), 305–323. https://doi.org/10/f9xt5j

Beretvas, S. N. (2011). Cross-classified and multiple-membership models. In J. J. Hox & J. K. Roberts (Series Ed.), *European Association for Methodology*. *Handbook of Advanced Multilevel Analysis* (pp. 313–334). New York, NY: Routledge.

Billsberry, J. (2014). The rise and rise of management edutainment. *Journal of Management Education, 38*(2), 151-159. https://doi.org/10.1177/1052562914524696

Blevins, D. P., Tsang, E. W. K., & Spain, S. M. (2015). Count-based research in management: Suggestions for improvement. *Organizational Research Methods*, *18*(1), 47–69. https://doi.org/10/gfz5h9

Brown, K. G. (2005). An Examination of the Structure and Nomological Network of Trainee Reactions: A Closer Look at "Smile Sheets". *Journal of Applied Psychology*, *90*(5), 991–1001. https://doi.org/10.1037/0021-9010.90.5.991

Brown, K. G., & Gerhardt, M. W. (2002). Formative evaluation: An integrative practice model and case study. *Personnel Psychology*, *55*(4), 951–983. https://doi.org/10/ft5zrf

Brown, K. G., & Sitzmann, T. (2011). Training and employee development for improved performance. In S. Zedeck (Ed.), *APA Handbook of Industrial and Organizational Psychology, Vol 2: Selecting and Developing Members for the Organization.* (pp. 469–503). https://doi.org/10.1037/12170-016

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review*, *6*(3), 263–296. https://doi.org/10.1177/1534484307303035

Burke, L. A., & Hutchins, H. M. (2008). A study of best practices in training transfer and proposed model of transfer. *Human Resource Development Quarterly, 19*, 107-128. https://doi.org/10.1002/hrdq.1230

Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I., & Mathieu, J. E. (1995). Toward theoretically based principles of training effectiveness: A model and initial empirical investigation. *Military Psychology*, *7*(3), 141–164. https://doi.org/10/cch45s

Casabianca, J. M., Lockwood, J. R., & McCaffrey, D. F. (2015). Trends in classroom observation scores. *Educational and Psychological Measurement*, *75*(2), 311–337. https://doi.org/10/gfz5h8

Chen, F., Brown, G., & Stokes, M. (2016). Fitting your favorite mixed models with PROC MCMC. *Proceedings of the SAS Global Forum 2016 Conference*. Presented at the Cary, NC. Cary, NC: SAS Institute, Inc.

Chen, Y.-J. (2001). Dimensions of transactional distance in the world wide web learning environment: A factor analysis. *British Journal of Educational Technology*, *32*(4), 459–470. https://doi.org/10/cj6nrg

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693. https://doi.org/10.1111/j.0956-7976.2004.00741.x

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, *85*(5), 678–707. https://doi.org/10/b368fq

Congdon, P. J., & McQueen, J. (2000). The stability of rater severity in large-scale assessment programs. *Journal of Educational Measurement*, *37*(2), 163–178. https://doi.org/10/b5h6hn

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives*, *8*, 1. https://doi.org/10/b2qf

Derouin, R. E., Fritzsche, B. A., & Salas, E. (2005). E-learning in organizations. *Journal of Management*, *31*(6), 920–940. https://doi.org/10/dtfk7r

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A content analysis of the content analysis literature in organization studies: research themes, data sources, and methodological refinements. *Organizational Research Methods*, *10*(1), 5–34. https://doi.org/10/dkkp3k

Eberly, M. B., Holley, E. C., Johnson, M. D., & Mitchell, T. R. (2011). Beyond internal and external: A dyadic theory of relational attributions. *Academy of Management Review*, *36*(4), 731–753. https://doi.org/10/bcsvbp

Eberly, M. B., Holley, E. C., Johnson, M. D., & Mitchell, T. R. (2017). It's not me, it's not you, it's us! An empirical examination of relational attributions. *Journal of Applied Psychology*, *102*(5), 711–731. https://doi.org/10/f98mvp

Ellis, L. (1994). *Research methods in the social sciences*. WCB Brown & Benchmark.

Evert, R. E., Payne, G. T., Moore, C. B., & McLeod, M. S. (2018). Top management team characteristics and organizational virtue orientation: An empirical examination of IPO firms. *Business Ethics Quarterly*, *28*(4), 427–461. https://doi.org/10.1017/beq.2018.3

Forgas, J. P., & George, J. M. (2001). Affective influences on judgments and behavior in organizations: An information processing perspective. *Organizational Behavior and Human Decision Processes, 86*(1), 3–34. https://doi.org/10.1006/obhd.2001.2971

Frenzel, A. C., Becker-Kurz, B., Pekrun, R., Goetz, T., & Lüdtke, O. (2018). Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *Journal of Educational Psychology*, *110*(5), 628–639. https://doi.org/10.1037/edu0000228

Gagné, R. M. (1962). Military training and principles of learning. *American Psychologist*, *17*(2), 83–91. https://doi.org/10/dzgzts

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*, 515–534. https://doi.org/10.1214/06-BA117A

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38. https://doi.org/10/f4k2h4

Glerum, D. R. (2011). *The Relationship Between Course Syllabi and Participant Evaluation Reactions Across Web-based and Face-to-Face Courses* [Master's thesis, University of Central Florida]. University of Central Florida Electronic Theses and Dissertations. http://purl.fcla.edu/fcla/etd/CFE0004113

Goldstein, I. L. (1993). *Training in Organizations*. Pacific Grove, CA: Brooks.

Goldstein, I. L., & Ford, J. K. (2002). *Training in Organizations: Needs Assessment, Development, and Evaluation*. Belmont, CA: Wadsworth/Thomson Learning.

Göllner, R., Wagner, W., Eccles, J. S., & Trautwein, U. (2018). Students' idiosyncratic perceptions of teaching quality in mathematics: A result of rater tendency alone or an expression of dyadic effects between students and teachers? *Journal of Educational Psychology*, *110*(5), 709–725. https://doi.org/10.1037/edu0000236

Gorham, J. (1988). The relationship between verbal teacher immediacy behaviors and student learning. *Communication Education*, *37*(1), 40–53. https://doi.org/10/bskbzh

Grossman, R., & Salas, E. (2011). The transfer of training: What really matters. *International Journal of Training and Development*, *15*(2), pp. 103-120. https://doi.org/10.1111/j.1468-2419.2011.00373.x

Gully, S., & Chen, G. (2010). Individual differences, attribute-treatment interactions, and training outcomes. In S. W. J. Kozlowski & E. Salas (Series Ed.), *SIOP Organizational Frontiers*. *Learning, Training, and Development in Organizations* (pp. 3-64). New York, NY: Routledge/Taylor & Francis.

Harik, P., Clauser, B. E., Grabovsky, I., Nungester, R. J., Swanson, D., & Nandakumar, R. (2009). An examination of rater drift within a generalizability theory framework. *Journal of Educational Measurement*, *46*(1), 43–58. https://doi.org/10/dt4gv2

Harman, R. P., Ellington, J. K., Surface, E. A., & Thompson, L. F. (2015). Exploring qualitative training reactions: Individual and contextual influences on trainee commenting. *Journal of Applied Psychology*, *100*(3), 894–916. https://doi.org/10/f7b9hw

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*, *89*(1), 92–102. https://doi.org/10/fh8gkk

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*(358), 320. https://doi.org/10/bpq49f

Holladay, C. L., & Quiñones, M. A. (2008). The influence of training focus and trainer characteristics on diversity training effectiveness. *Academy of Management Learning & Education, 7*(3), 343–354. https://doi.org/10.5465/AMLE.2008.34251672

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Addison-Wesley.

Howardson, G. N., & Behrend, T. S. (2016). Coming full circle with reactions: Understanding the structure and correlates of trainee reactions through the affect circumplex. *Academy*

of Management Learning & Education, 15(3), 471–492. https://doi.org/10.5465/amle.2014.0012

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications*. New York, NY: Routledge.

Kaman, V. S. (1985). Predict training fortunes. *Personnel Journal*, *64*, 42–47.

Keller, M. M., Hoy, A. W., Goetz, T., & Frenzel, A. C. (2016). Teacher enthusiasm: reviewing and redefining a complex construct. *Educational Psychology Review*, *28*(4), 743–769. https://doi.org/10.1007/s10648-015-9354-y

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal Relationships*. New York, NY: Wiley.

Kim, L. E., Dar-Nimrod, I., & MacCann, C. (2018). Teacher personality and teacher effectiveness in secondary school: personality predicts teacher support and student self-efficacy but not academic achievement. *Journal of Educational Psychology*, *110*(3), 309–323. https://doi.org/10.1037/edu0000217

Kirkpatrick, D. L. (1956). How to start an objective evaluation of your training program. *Journal of the American Society of Training Directors*, *10*, 18–22.

Kirkpatrick, D. L. (1996). Great ideas revisited: Revisiting Kirkpatrick's four-level-model. *Training and Development*, *50*, 54–59.

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, *86*(3), 643–680. https://doi.org/10/gfj8hf

Kozlowski, S. W. J., Chao, G. T., & Jensen, J. M. (2010). Building an infrastructure for organizational learning: A multilevel approach. In S. W. J. Kozlowski & E. Salas (Series

Ed.), *SIOP Organizational Frontiers*. *Learning, Training, and Development in Organizations* (pp. 363–403). New York, NY: Routledge/Taylor & Francis.

Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Series Ed.), *SIOP Lessons for Practice*. *Creating, Implementing, and Managing Effective Training and Development: State-of-the-art Lessons for Practice* (pp. 331–376). San Francisco, CA: Jossey-Bass.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*(2), 311–328. https://doi.org/10/fk7nhq

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*(4), 722–752. https://doi.org/10/f392tm

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, *36*, 143–152. https://doi.org/10/f5fkp4

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. T. Reis (Ed.), *New Directions for Methodology of Social and Behavioral Science* (Vol. 15, pp. 41–56). San Francisco, CA: Jossey-Bass.

Liu, W., & Cela, J. (2008). Count data models in SAS. *SAS Global Forum*, *317*, 1–12.

LoPilato, A. C., Carter, N. T., & Wang, M. (2015). Updating generalizability theory in management research: Bayesian estimation of variance components. *Journal of Management*, *41*(2), 692–717. https://doi.org/10/gcp2sf

Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness. *Academy of Management Journal, 35*(4), pp. 828-847. https://doi.org/10.2307/256317

Mathieu, J. E., & Tesluk, P. E. (2010). A multilevel perspective on training and development effectiveness. In S. W. J. Kozlowski & E. Salas (Series Ed.), *SIOP Organizational Frontiers*. *Learning, Training, and Development in Organizations* (pp. 405–440). New York, NY: Routledge/Taylor & Francis.

McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. (2018). What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management*, *44*(7), 2909–2933. https://doi.org/10/gcwbkm

McKenny, A. F., Short, J. C., & Newman, S. M. (2012). CAT Scanner (Version 1.0) [Software]. Available from http://www.catscanner.net/

McKenny, A. F., Short, J. C., & Payne, G. T. (2013). Using computer-aided text analysis to elevate constructs: An illustration using psychological capital. *Organizational Research Methods*, *16*(1), 152–184. https://doi.org/10.1177/1094428112459910

Mehrabian, A. (1981). *Silent Messages: Implicit Communication of Emotions and Attitudes* (2nd Ed.). Belmont, CA: Wadsworth.

Meyer, R. H. (2001). *Estimation of Teacher and School Performance in the Denver Public Schools: A Feasibility Study*. Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Educational Research.

Meyers, L. E. (2012). Cross-classified random effects models in institutional research. *New Directions for Institutional Research*, *2012*(154), 77–93. https://doi.org/10.1002/ir.20015

Moore, M. G. (1989). Three types of interaction. *American Journal of Distance Education*, *3*(2), 1–7. https://doi.org/10/bnphbg

Moore, M. G. (1997). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical Principles of Distance Education* (pp. 22-38). New York: Routledge.

Moreno, R., Reislein, M., & Ozogul, G. (2010). Using virtual peers to guide visual attention during learning: A test of the persona hypothesis. *Journal of Media Psychology*, *22*(2), 52–60. https://doi.org/10.1027/1864-1105/a000008

Morris, M. (1984). The evaluation of training. *Industrial and Commercial Training*, *16*(2), pp. 9-16. https://doi.org/10.1108/eb003971

Noe, R.A. (1986), Trainees' Attributes and Attitudes: Neglected Influences on Training Effectiveness. *Academy of Management Review*, *11*(4), 736–749. https://doi.org/10.5465/amr.1986.4283922

Noe, R. A., Clarke, A. D. M., & Klein, H. J. (2014). Learning in the twenty-first-century workplace. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 245–275. https://doi.org/10/gckf46

Odden, A., Borman, G., & Fermanich, M. (2004). Assessing teacher, classroom, and school effects, including fiscal effects. *Peabody Journal of Education*, *79*(4), 4–32. https://doi.org/10/bmv8db

Offir, B., Lev, Y., Lev, Y., Barth, I., & Shteinbok, A. (2004). An integrated analysis of verbal and nonverbal interaction in conventional and distance learning environments. *Journal of Educational Computing Research*, *31*(2), 101–118. https://doi.org/10/bbvz5f

Okan, Z. (2003). Edutainment: Is learning at risk? *British Journal of Educational Technology, 34*(3), 255-264. https://doi.org/10.1111/1467-8535.00325

Osborn, C. (2018, October 26). What makes a good trainer? [Blog]. Retrieved from The Training Associates website: https://thetrainingassociates.com/blog/what-makes-good-trainer/

Ouwehand, K., van Gog, T., & Paas, F. (2015). Designing effective video-based modeling examples using gaze and gesture cues. *Educational Technology & Society*, *18*(4), 78–88.

Patel, L. (2010). *State of the industry report: ASTD's definitive review of workplace learning and development trends*. Alexandria, VA: American Society for Training & Development.

Pekrun, R. (2006). The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC 2007*. The University of Texas at Austin. https://www.liwc.net/LIWC2007LanguageManual.pdf

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile?: An analytical framework for answering the question. *Quality Assurance in Education*, *15*(2), 178–191. https://doi.org/10/fncfx8

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rocklage, M. D., & Fazio, R. H. (2016). On the dominance of attitude emotionality. *Personality and Social Psychology Bulletin*, *42*(2), 259–270. https://doi.org/10.1177/0146167215623273

Rocklage, M. D., & Fazio, R. H. (2018). Attitude accessibility as a function of emotionality. *Personality and Social Psychology Bulletin*, *44*(4), 508–520. https://doi.org/10.1177/0146167217743762

Rodale, J. I. (1978). *The synonym finder*. Rodale.

Roorda, D. L., Jak, S., Zee, M., Oort, F. J., & Koomen, H. M. Y. (2017). Affective teacher–student relationships and students' engagement and achievement: A meta-analytic update and test of the mediating role of engagement. *School Psychology Review*, *46*(3), 239–261. https://doi.org/10/gfz5h4

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, *81*(4), 493–529. https://doi.org/10/dgm6bn

Ruona, W. E. A., Leimbach, M., Holton, E. F., & Bates, R. (2002). The relationship between learner utility reactions and predicted learning transfer among trainees. *International Journal of Training and Development*, *6*(4), 218–228. https://doi.org/10/fg2v23

Rusbult, C. E., & Van Lange, P. A. M. (2008). Why we need interdependence theory. *Social and Personality Psychology Compass*, *2*(5), 2049–2070. https://doi.org/10/fkwk66

Sanders, W. L. (2000). Value-added assessment from student achievement data. *Journal of Personnel Evaluation in Education*, *14*(4), 329–339. https://doi.org/10/dchmgj

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*. Knoxville, TN: University of Tennessee, Value-Added Research and Assessment Center.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110. https://doi.org/10/c7k4kg

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior, Vol. 2* (p. 527–561). Guilford.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454–499. https://doi.org/10/bb73d5

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods*, *13*(2), 320–347. https://doi.org/10.1177/1094428109335949

Sitzmann, T., Brown, K. G., Casper, W. J., Ely, K., & Zimmerman, R. D. (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology*, *93*(2), 280–295. https://doi.org/10.1037/0021-9010.93.2.280

Sitzmann, T., & Johnson, S. (2014). The paradox of seduction by irrelevant details: How irrelevant information helps and hinders self-regulated learning. *Learning and Individual Differences*, *34*, 1–11. https://doi.org/10/gfz5jf

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: a meta-analysis. *Personnel Psychology*, *59*(3), 623–664. https://doi.org/10/dmn926

Sitzmann, T., & Weinhardt, J. M. (2018). Training engagement theory: A multilevel perspective on the effectiveness of work-related training. *Journal of Management*, *44*(2), 732–756. https://doi.org/10/gft8nx

Steiner, D. D., Dobbins, G. H., & Trahan, W. A. (1991). The trainer-trainee interaction: An attributional model of training. *Journal of Organizational Behavior*, *12*(4), 271–286. https://doi.org/10/fh3x4k

Tannenbaum, S. I., Cannon-Bowers, J. A., Salas, E., & Mathieu, J. E. (1993). *Factors that influence training effectiveness: A conceptual model and longitudinal analysis* (No. 93–011). Orlando, FL: Naval Air Warfare Center Training Systems Division.

Thayer, P. W., & McGehee, W. (1977). "On the effectiveness of not holding a formal training course," *Personnel Psychology*, 30, 455-456. https://doi.org/10.1111/j.1744-6570.1977.tb00438.x

Thomas, G., & Thorpe, S. (2019). Enhancing the facilitation of online groups in higher education: a review of the literature on face-to-face and online group-facilitation. *Interactive Learning Environments*, *27*(1), 62–71. https://doi.org/10.1080/10494820.2018.1451897

Towler, A. J., & Dipboye, R. L. (2001). Effects of trainer expressiveness, organization, and trainee goal orientation on training outcomes. *Journal of Applied Psychology*, *86*(4), 664–673. https://doi.org/10/b76qr2

Tracey, J. B., Hinkin, T. R., Tannenbaum, S., & Mathieu, J. E. (2001). The influence of individual characteristics and the work environment on varying levels of training outcomes. *Human Resource Development Quarterly*, *12*(1), 5–23. https://doi.org/10/cxw73f

Trougakos, J. P., Beal, D. J., Green, S. G., & Weiss, H. M. (2008). Making the break count: An episodic examination of recovery activities, emotional experiences, and positive affective

displays. *Academy of Management Journal*, *51*(1), 131-146. https://doi.org/10.5465/AMJ.2008.30764063

Tzabbar, D., Tzafrir, S., & Baruch, Y. (2017). A bridge over troubled water: Replication, integration and extension of the relationship between HRM practices and organizational performance using moderating meta-analysis. *Human Resource Management Review*, *27*(1), 134–148. https://doi.org/10/gfz5jd

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, *108*(5), 705-721. http://doi.org/10.1037/edu0000075

Warr, P., Allan, C., & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology*, *72*, 351-375. https://doi.org/10.1348/096317999166725

Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology, 48*(2), 347–375. https://doi.org/10.1111/j.1744-6570.1995.tb01761.x

Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). 4 reliability coefficients and generalizability theory. In *Handbook of Statistics* (Vol. 26, pp. 81–124). https://doi.org/10.1016/S0169-7161(06)26004-8

Webster, J., & Martocchio, J. J. (1995). The differential effects of software training previews on training outcomes. *Journal of Management, 21*(4), 757–787. https://doi.org/10.1177/014920639502100409

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 209–231). New York, NY: Guilford Press.

Wilson, M., & Case, H. (2000). An examination of variation in rater severity over time: A study of rater drift. In M. Wilson & G. Engelhard (Eds.), *Objective Measurement: Theory into Practice* (Vol. 5, pp. 113–134). Stamford, CT: Ablex.

Wulfeck II., W. H., Ellis, J. A., Richards, R. E., Wood, N. D., & Merrill, M. D. (1978). *The Instructional Quality Inventory: I. Introduction and Overview* (No. NPRDC SR 79-3). San Diego, CA: Navy Personnel Research and Development Center.

Zachary, M. A., McKenny, A., Short, J. C., & Payne, G. T. (2011). Family business and market orientation: construct validation and comparative analysis. *Family Business Review*, *24*(3), 233–251. https://doi.org/10.1177/0894486510396871

Table 1

Means, Standard Deviations, and Item-Level and Scale-Level Correlations among Trainee Reactions in Studies 1 and 2

Variable	S1: *M* (*SD*)	S1 α	1	2	3	4	5	6	7	8	9	10	11	12	S2: *M* (*SD*)	S2 α
1. Overall Trainee Reactions Scale	4.54 (0.64)	.94	--	.84	.91	.88	.81	.83	.83	.83		.82	.79	.79	4.57 (0.53)	.92
2. Overall Trainee Reactions (Single Item)				--	.67	.84	.58	.61	.60	.63		.65	.55	.56	4.52 (0.70)	
3. Satisfaction Scale	4.72 (0.65)	.96	.92		--	.62	.89	.91	.92	.88		.86	.89	.88	4.76 (0.54)	.95
4. Utility Scale	4.36 (0.73)	.89	.93		.71	--	.54	.56	.54	.59		.61	.49	.51	4.31 (0.65)	.83
5. Knowledge of Subject[s]	4.74 (0.66)		.85		.92	.66	--	.84	.81	.77		.69	.73	.74	4.82 (0.52)	
6. Preparation / Organization[s]	4.71 (0.69)		.85		.93	.66	.87	--	.80	.77		.73	.80	.75	4.78 (0.58)	
7. Creation of a Positive Learning Atmosphere[s]	4.74 (0.70)		.85		.93	.66	.80	.81	--	.82		.72	.78	.82	4.78 (0.58)	
8. Equal Concern for Theory and Practice[s]	4.72 (0.68)		.87		.94	.68	.84	.83	.87	--		.69	.73	.73	4.72 (0.61)	
9. Effective Classroom Management[s]	4.69 (0.76)		.85		.94	.66	.81	.83	.86	.86	--					
10. Email Interaction with Instructor[s]												--	.74	.71	4.63 (0.72)	
11. Accessibility / Responsiveness[s]													--	.78	4.77 (0.64)	
12. Warmth / Rapport[s]														--	4.80 (0.57)	
13. Comparison to Prior Training Courses[u]	4.43 (0.85)		.87		.73	.88	.68	.68	.68	.70	.69				4.43 (0.76)	
14. Usefulness of Text[u]	4.12 (1.01)		.67		.43	.80	.40	.40	.39	.41	.39				4.23 (0.93)	
15. Usefulness of Workbook[u]	4.28 (0.87)		.73		.51	.83	.48	.47	.47	.48	.47					
16. Usefulness of Experiential Activities[u]	4.50 (0.81)		.86		.73	.85	.67	.67	.69	.70	.69					
17. Usefulness of Website Materials[u]															4.43 (0.75)	
18. Usefulness of Discussion Forum[u]															3.95 (1.01)	
19. Classroom Application[u]	4.45 (0.83)		.77		.60	.81	.56	.56	.55	.58	.55				4.50 (0.75)	

Note. Study 1 (*N* = 7,097) statistics are presented below the diagonal; Study 2 (*N* = 3,182) statistics are presented above the diagonal. *M* = mean; *SD* = standard deviation.

[s] = item from satisfaction scale. [u] = item from utility scale.

All correlations significant at $p < .05$, two-tailed.

Table 1 (Continued)

Means, Standard Deviations, and Item-Level and Scale-Level Correlations among Trainee Reactions in Studies 1 and 2

Variable	13	14	15	16	17	18	19
1. Overall Trainee Reactions Scale	.81	.62			.72	.63	.68
2. Overall Trainee Reactions (Single Item)	.84	.59			.66	.53	.69
3. Satisfaction Scale	.62	.38			.52	.41	.50
4. Utility Scale	.84	.76			.79	.74	.75
5. Knowledge of Subject[s]	.54	.35			.47	.34	.44
6. Preparation / Organization[s]	.57	.35			.48	.36	.45
7. Creation of a Positive Learning Atmosphere[s]	.56	.33			.45	.36	.44
8. Equal Concern for Theory and Practice[s]	.59	.36			.51	.40	.47
9. Effective Classroom Management[s]							
10. Email Interaction with Instructor[s]	.59	.36			.50	.45	.48
11. Accessibility / Responsiveness[s]	.50	.30			.41	.32	.40
12. Warmth / Rapport[s]	.52	.30			.42	.33	.42
13. Comparison to Prior Training Courses[u]	--	.58			.62	.51	.63
14. Usefulness of Text[u]	.56	--			.51	.37	.46
15. Usefulness of Workbook[u]	.62	.67	--				
16. Usefulness of Experiential Activities[u]	.80	.51	.61	--			
17. Usefulness of Website Materials[u]					--	.51	.49
18. Usefulness of Discussion Forum[u]						--	.40
19. Classroom Application[u]	.70	.50	.55	.69			--

Note. Study 1 (N = 7,097) statistics are presented below the diagonal; Study 2 (N = 3,182) statistics are presented above the diagonal. M = mean; SD = standard deviation.

[s] = item from satisfaction scale. [u] = item from utility scale.

All correlations significant at $p < .05$, two-tailed.

Table 2

Confirmatory Factor Analyses (CFA) of Online and Face-to-Face Trainee Reaction Items from Studies 1 and 2

	Factor Loadings	
Item	Face-to-Face	Online
Knowledge of Subject (Satisfaction)	.90	.88
Preparation / Organization (Satisfaction)	.90	.90
Creation of a Positive Learning Atmosphere (Satisfaction)	.91	.91
Equal Concern for Theory and Practice (Satisfaction)	.93	.87
Effective Classroom Management (Satisfaction)	.92	-
Email Interaction with Instructor (Satisfaction)	-	.81
Accessibility / Responsiveness (Satisfaction)	-	.87
Warmth / Rapport (Satisfaction)	-	.86
Comparison with prior Training Courses (Utility)	.90	.88
Usefulness of Text (Utility)	.63	.65
Classroom Application (Utility)	.77	.70
Usefulness of Workbook (Utility)	.70	-
Usefulness of Experiential Activities (Utility)	.89	-
Usefulness of Website Materials (Utility)	-	.73
Usefulness of Discussion Forum (Utility)	-	.60
Fit indices		
$\chi 2$ (df)	2858.89 (34)	1349.47 (53)
RMSEA, [90% CI]	.11, [.11, .11]	.09, [.08, .09]
p-close	$p < .01$	$p < .01$
SRMR	.04	.04
GFI	.92	.94
TLI (NNFI)	.95	.95
CFI	.96	.96

Note. Completely standardized solutions. Hyphens indicate items not present in either online or face-to-face course. RMSEA = Root Mean Square of Approximation; SRMR = Standardized Root Mean Square Residual; GFI=Goodness of Fit Index; TLI = Tucker—Lewis Index; NNFI = Non-normed Fit Index; CFI = Comparative Fit Index; df = Degrees of Freedom; CI = Confidence Interval.

Table 3

Unconditional Models of the Trainee Reactions Surveys from Study 1: Face-to-Face Courses

Trainee Reactions	Random Effect	Variance Component	SE	ICC
Overall Reactions (Aggregated)	Trainer	0.135*	0.035	0.296
	Content	0.007	0.005	0.016
	Trainer*Content	0.060*	0.009	0.132
	Residual	0.253*	0.004	
Satisfaction	Trainer	0.153*	0.038	0.318
	Content	0.002	0.004	0.005
	Trainer*Content	0.055*	0.008	0.114
	Residual	0.270*	0.005	
Knowledge of Subject	Trainer	0.142*	0.035	0.295
	Content	0.002	0.004	0.005
	Trainer*Content	0.051*	0.008	0.105
	Residual	0.287*	0.005	
Preparation / Organization	Trainer	0.140*	0.036	0.266
	Content	0.003	0.004	0.006
	Trainer*Content	0.060*	0.009	0.114
	Residual	0.323*	0.006	
Creation of a Positive Learning Atmosphere	Trainer	0.169*	0.042	0.305
	Content	0.003	0.004	0.006
	Trainer*Content	0.054*	0.009	0.098
	Residual	0.327*	0.006	
Equal Concern for Theory and Practice	Trainer	0.139*	0.035	0.268
	Content	0.001	0.003	0.002
	Trainer*Content	0.048*	0.008	0.092
	Residual	0.331*	0.006	
Effective Classroom Management	Trainer	0.198*	0.049	0.304
	Content	0.003	0.005	0.004
	Trainer*Content	0.065*	0.011	0.100
	Residual	0.385*	0007	
Utility	Trainer	0.119*	0.033	0.208
	Content	0.022*	0.010	0.039
	Trainer*Content	0.073*	0.011	0.127
	Residual	0.358*	0.006	
Text Usefulness	Trainer	0.070*	0.030	0.066
	Content	0.157*	0.047	0.148
	Trainer*Content	0.124*	0.020	0.117
	Residual	0.708*	0.012	
Workbook Usefulness	Trainer	0.065*	0.025	0.082
	Content	0.038*	0.016	0.048
	Trainer*Content	0.093*	0.016	0.118
	Residual	0.595*	0.010	
Experiential Activities Usefulness	Trainer	0.180*	0.046	0.248
	Content	0.001	0.005	0.001
	Trainer*Content	0.074*	0.012	0.102
	Residual	0.470*	0.008	
Class Application	Trainer	0.121*	0.033	0.168
	Content	0.012	0.008	0.016
	Trainer*Content	0.061*	0.011	0.084
	Residual	0.528*	0.009	
Course Comparison	Trainer	0.183*	0.048	0.233

Trainee Reactions	Random Effect	Variance Component	SE	ICC
	Content	0.003	0.007	0.004
	Trainer*Content	0.096*	0.015	0.122
	Residual	0.505*	0.009	

Note. Random Effects based on unconditional model. Satisfaction, Utility, and Overall Reactions are averages of the relevant items from the trainee reactions subscales and are presented in bold. SE = Standard Error of Variance Component; ICC = Intraclass Correlation Coefficient.
*$p < .05$, two-tailed.

Table 4

Unconditional Models of the Computer-Aided Text Analysis in Study 1: Face-to-Face Course

Item	Trainee Reaction Dimension	Random Effect	Modal Variance Component	2.5% CrI (Low)	97.5% CrI (High)	ICC
Open-ended Satisfaction	**Satisfaction**	Trainer	0.001	<0.001	0.004	0.315
		Content	<0.001	<0.001	0.004	0.001
		Trainer*Content	0.001	<0.001	0.004	0.682
		Residual	<0.001	<0.001	0.002	
	Utility	Trainer	0.004	<0.001	0.045	0.111
		Content	<0.001	<0.001	0.063	0.001
		Trainer*Content	0.030	<0.001	0.047	0.832
		Residual	0.002	<0.001	0.004	
Open-ended Utility	**Satisfaction**	Trainer	<0.001	<0.001	0.014	0.034
		Content	<0.001	<0.001	0.004	0.004
		Trainer*Content	0.001	<0.001	0.012	0.937
		Residual	<0.001	<0.001	0.007	
	Utility	Trainer	<0.001	<0.001	0.015	<0.001
		Content	0.005	<0.001	0.025	0.872
		Trainer*Content	<0.001	<0.001	0.011	0.018
		Residual	0.001	<0.001	0.008	

Note. Variance components based on unconditional model with an over-dispersion (residual) parameter and scaled with total words per comment as an offset variable (Hox, 2010; McCullagh & Nelder, 1989). $k_{posterior}$ = 10,000; ICC = Intraclass Correlation Coefficient; CrI = Credibility Interval.

Table 5

Unconditional Models of the Trainee Reactions Surveys in Study 2: Online Course

Trainee Reactions	Random Effect	Variance Component	SE	ICC
Global Reactions (Single Item)	Trainer	0.005	0.006	0.011
	Content	0.013*	0.007	0.026
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.474*	0.012	
Overall Reactions (Aggregated)	Trainer	0.009*	0.005	0.033
	Content	0.008*	0.004	0.029
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.260*	0.007	
Satisfaction	Trainer	0.019*	0.008	0.066
	Content	0.005	0.003	0.017
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.267*	0.007	
Knowledge of Subject	Trainer	0.010*	0.004	0.037
	Content	0.001	0.002	0.003
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.258*	0.007	
Preparation / Organization	Trainer	0.020*	0.009	0.057
	Content	0.005	0.004	0.015
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.317*	0.008	
Creation of a Positive Learning Atmosphere	Trainer	0.014*	0.007	0.042
	Content	0.006	0.004	0.018
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.319*	0.008	
Equal Concern for Theory and Practice	Trainer	0.007	0.005	0.019
	Content	0.007*	0.004	0.020
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.363*	0.009	
Accessibility / Responsiveness	Trainer	0.052*	0.020	0.125
	Content	0.006	0.005	0.014
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.359*	0.009	
Email Interaction with Instructor	Trainer	0.035*	0.013	0.066
	Content	0.010*	0.006	0.020
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.479*	0.012	
Warmth / Rapport	Trainer	0.023*	0.008	0.070
	Content	0.001	0.002	0.003
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.307*	0.008	
Utility	Trainer	0.003	0.004	0.006
	Content	0.015*	0.007	0.036
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.405*	0.010	
Text Usefulness	Trainer	<0.001	<0.001	<0.001
	Content	0.072*	0.020	0.084
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.785*	0.020	
Website Materials Usefulness	Trainer	0.005	0.005	0.009

Trainee Reactions	Random Effect	Variance Component	SE	ICC
	Content	<0.001	<0.001	<0.001
	Trainer*Content	0.008	0.005	0.015
	Residual	0.549*	0.014	
Forum Usefulness	Trainer	0.004	0.008	0.004
	Content	0.053*	0.021	0.051
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.967*	0.024	
Class Application	Trainer	<0.001	<0.001	<0.001
	Content	0.016*	0.006	0.028
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.552*	0.014	
Course Comparison	Trainer	0.008	0.006	0.014
	Content	0.012*	0.007	0.021
	Trainer*Content	<0.001	<0.001	<0.001
	Residual	0.553*	0.014	

Note. Random Effects based on unconditional model. Satisfaction, Utility, and Overall Reactions are averages of the relevant items from the trainee reactions subscales and are presented in bold. SE = Standard Error of Variance Component; ICC = Intraclass Correlation Coefficient.
*$p < .05$, two-tailed.

Table 6

Unconditional Models of the Computer-Aided Text Analysis in Study 2: Online Course

Item	Trainee Reaction Dimension	Random Effect	Modal Variance Component	2.5% CrI (Low)	97.5% CrI (High)	ICC
Open-Ended Satisfaction	**Satisfaction**	Trainer	0.001	<0.001	0.017	0.838
		Content	<0.001	<0.001	0.016	0.144
		Trainer*Content	<0.001	<0.001	0.014	0.001
		Residual	<0.001	<0.001	0.009	
	Utility	Trainer	<0.001	<0.001	0.044	0.001
		Content	<0.001	<0.001	0.050	0.004
		Trainer*Content	0.021	<0.001	0.049	0.940
		Residual	0.001	<0.001	0.003	
Overall Course: Targeted	**Satisfaction**	Trainer	0.015	<0.001	0.028	0.976
		Content	<0.001	<0.001	0.021	<0.001
		Trainer*Content	<0.001	<0.001	0.030	0.023
		Residual	<0.001	<0.001	0.014	
	Utility	Trainer	0.004	<0.001	0.033	0.093
		Content	<0.001	<0.001	0.040	0.001
		Trainer*Content	<0.001	<0.001	0.050	<0.001
		Residual	0.038	<0.001	0.030	
Overall Course: Miscellaneous	**Satisfaction**	Trainer	<0.001	<0.001	0.100	0.003
		Content	<0.001	<0.001	0.094	0.003
		Trainer*Content	0.001	<0.001	0.106	0.928
		Residual	<0.001	<0.001	0.089	
	Utility	Trainer	0.003	<0.001	0.143	0.036
		Content	<0.001	<0.001	0.160	<0.001
		Trainer*Content	0.078	<0.001	0.179	0.952
		Residual	0.001	<0.001	0.406	
Personal Results	**Satisfaction**	Trainer	<0.001	<0.001	0.037	0.047
		Content	<0.001	<0.001	0.032	0.682
		Trainer*Content	<0.001	<0.001	0.028	0.264
		Residual	<0.001	<0.001	0.014	
	Utility	Trainer	0.001	<0.001	0.008	0.137
		Content	<0.001	<0.001	0.010	<0.001
		Trainer*Content	0.002	<0.001	0.009	0.421
		Residual	0.002	<0.001	0.008	
Change Recommend.	**Satisfaction**	Trainer	<0.001	<0.001	0.062	0.002
		Content	<0.001	<0.001	0.056	0.008
		Trainer*Content	0.020	<0.001	0.057	0.989
		Residual	<0.001	<0.001	0.045	
	Utility	Trainer	<0.001	<0.001	0.247	<0.001

		Content	<0.001	<0.001	0.119	0.001
		Trainer*Content	0.299	<0.001	0.268	0.988
		Residual	0.003	<0.001	0.123	
Strengths	**Satisfaction**	Trainer	<0.001	<0.001	0.045	<0.001
		Content	0.001	<0.001	0.034	0.027
		Trainer*Content	0.030	<0.001	0.033	0.973
		Residual	<0.001	<0.001	0.019	
	Utility	Trainer	0.007	<0.001	0.082	0.824
		Content	<0.001	<0.001	0.060	0.052
		Trainer*Content	0.001	<0.001	0.062	0.108
		Residual	<0.001	<0.001	0.017	
Weaknesses	**Satisfaction**	Trainer	0.035	<0.001	0.133	0.993
		Content	<0.001	<0.001	0.051	0.004
		Trainer*Content	<0.001	<0.001	0.066	0.001
		Residual	<0.001	<0.001	0.042	
	Utility	Trainer	0.019	<0.001	0.109	0.891
		Content	0.001	<0.001	0.017	0.041
		Trainer*Content	<0.001	<0.001	0.079	<0.001
		Residual	0.001	<0.001	0.067	

Note. Variance components based on unconditional model with an over-dispersion (residual) parameter and scaled with total words per comment as an offset variable (Hox, 2010; McCullagh & Nelder, 1989). $k_{posterior}$ = 10,000; ICC = Intraclass Correlation Coefficient; CrI = Credibility Interval.

Table 7

Overview of Trainee Reactions Computer-Aided Text Analysis (CATA) Results Across Studies 1 and 2

Trainee Reaction Dimension	Study 1: Face-to-Face		Study 2: Online						
	Satisfaction	Utility	Satisfaction	Overall - Targeted	Overall - Misc.	Results Experienced	Change Recommendations	Strengths	Weaknesses
Satisfaction	Trainer	Equal	Trainer	Trainer	Equal	Content	Equal	Equal	Trainer
Utility	Trainer	Content	Equal	Trainer	Equal	Trainer	Equal	Trainer	Trainer

Note: Trainer is the largest source of variance in 9 or 50.00% of analyses, Course content is the largest source of variance in 2 or 11.11% of analyses, Trainer and Course content are relatively equal sources of variance in 7 or 38.89% of analyses; The largest source of variance determination was based on whether there was a >=5% difference in ICC size across both sources of variance (a conservative cutoff value).

Table 8

Frequencies and Percentages of Primary Variance Sources by Trainee Reaction Item-Level Measures across Studies 1 & 2

Trainee Reaction Dimensions & Measures	Face-to-Face (Study 1)			Online (Study 2)			Overall (Studies 1-2)		
	Trainer	Content	Equal	Trainer	Content	Equal	Trainer	Content	Equal
Overall									
Global (Item)	--	--	--	0 (0.00%)	0 (0.00%)	1 (100.00%)	0 (0.00%)	0 (0.00%)	1 (100.00%)
TOTALS	--	--	--	0 (0.00%)	0 (0.00%)	1 (100.00%)	0 (0.00%)	0 (0.00%)	1 (100.00%)
Satisfaction									
Likert-type	5 (100.00%)	0 (0.00%)	0 (0.00%)	2 (28.57%)	0 (0.00%)	5 (71.43%)	7 (58.33%)	0 (0.00%)	5 (41.67%)
CATA	1 (50.00%)	0 (0.00%)	1 (50.00%)	3 (42.86%)	1 (14.29%)	3 (42.86%)	4 (44.44%)	1 (11.11%)	4 (44.44%)
TOTALS	6 (85.71%)	0 (0.00%)	1 (14.29%)	5 (35.71%)	1 (7.14%)	8 (57.14%)	11 (52.38%)	1 (4.76%)	9 (42.86%)
Utility									
Likert-type	4 (80.00%)	1 (20.00%)	0 (0.00%)	0 (0.00%)	1 (20.00%)	4 (80.00%)	4 (40.00%)	2 (20.00%)	4 (40.00%)
CATA	1 (50.00%)	1 (50.00%)	0 (0.00%)	4 (57.14%)	0 (0.00%)	3 (42.86%)	5 (55.56%)	1 (11.11%)	3 (33.33%)
TOTALS	5 (71.43%)	2 (28.57%)	0 (0.00%)	4 (33.33%)	1 (8.33%)	7 (58.33%)	9 (50.00%)	2 (11.11%)	7 (38.89%)
TOTALS (out of all reactions)	**11 (78.57%)**	**2 (14.29%)**	**1 (7.14%)**	**9 (33.33%)**	**2 (7.41%)**	**16 (59.26%)**	**20 (50.00%)**	**3 (7.50%)**	**17 (42.50%)**

Note: Trainer is the largest source of variance in 20 or 50.00% of analyses, course content is the largest source of variance in 3 or 7.50% of analyses, trainer and course content are relatively equal sources of variance in 17 or 42.50% of analyses. The largest source of variance determination was made based on whether there was a >=5% difference in ICC size across both sources of variance (a conservative cutoff value). Overall reactions (aggregated) were not included, given the component items are already included in their respective Satisfaction and Utility Likert-type rows.

Table 9

Summary of Hypotheses and Support Across Studies 1 and 2

Hypotheses	Study 1: Face-to-Face Courses	Study 2: Online Courses
Hypothesis 1*:* Trainee reactions reflect more between-trainer variance than between-content variance.	**SUPPORTED**	**PARTIALLY SUPPORTED**
Hypothesis 2*:* Trainee satisfaction reactions reflect more between-trainer variance than between-content variance.	**SUPPORTED**	**PARTIALLY SUPPORTED**
Hypothesis 3*:* Trainee utility reactions reflect more between-content variance than between-trainer variance.	**NOT SUPPORTED**	**NOT SUPPORTED**

Note. Support was obtained if at least 50.00% of items reflected the hypothesized variance source. Otherwise, partial support was obtained if most of the items reflected the hypothesized variance source, although not exceeding 50.00%. This value was chosen as a conservative criterion given that it was possible for measures to reflect content and trainer variance to a similar degree.

Supplemental Material

The Development and Validation of a CATA Measure of Trainee Reactions

To develop and validate measures of satisfaction and utility, we followed the two-phase dictionary development process (Short, Broberg, Cogliser, & Brigham, 2010). In the first phase, words plausibly reflective of each construct are deductively identified from existing instruments and from synonyms of words thought to be closely associated with the construct. For example, all synonyms associated with "enjoy," "enjoyable," "enjoyment," "satisfaction," "satisfactory," "satisfied," and "satisfy" were identified in Rodale's (1978) The Synonym Finder and Thesaurus.com to provide an initial deductive list for our satisfaction dictionary. After ensuring that all forms of the identified synonym lexemes were included in the list, the initial deductive lists for satisfaction contained 1,327 words and for utility contained 703 words.

To maximize dictionary content validity, the second phase calls for researchers to draw words from the sampled texts to identify potentially relevant words in use by participants that were not captured in the initial deductive word lists (Short et al., 2010). This also improves content validity by introducing typographical errors used in the sample of texts for consideration. To build this inductive list, we identified every word used in any of the trainee reaction comments. To mitigate the risk of coder fatigue in the final word list evaluation process, one author eliminated a number of words that were clearly not relevant to training evaluation (e.g., "and," "meat," "many") or were duplicative of words already identified in the deductive process, resulting in an initial inductive word list of 694 words for satisfaction and 712 words for utility.

After combining the inductive and deductive lists, the authors manually identified several additional words that were thought to be indicative of each construct but were not identified by the deductive-inductive process (e.g., "nicely done" for satisfaction). This uncovered an

additional 76 words for satisfaction and 139 words for utility. The resulting initial word lists for evaluation consisted of 2,097 words for satisfaction and 1,554 words for utility.

Three authors independently evaluated the word lists in their entirety to assess whether each word would be reflective of the construct if found in training evaluation texts (Short et al., 2010). We used Holsti's (1969) measure adapted to accommodate three judges to assess the interrater reliability (e.g., Zachary, McKenny, Short, & Payne, 2013). We found interrater reliability of 0.77 for satisfaction and 0.84 for utility, comparing favorably to guidance indicating that coefficients greater than 0.75 be interpreted as acceptable (e.g., Ellis, 1994). The judges then resolved all disagreements through discussion, resulting in final dictionaries of 531 words for satisfaction and 272 words for utility. These dictionaries are presented at the end of the appendix and are provided in CAT Scanner dictionary format as supplemental files on APA.org and CATScanner.net.

We used the CAT Scanner software to measure our constructs (McKenny, Short, & Newman, 2012). CAT Scanner was developed by management researchers as a free alternative to commercial CATA packages and has been used in the organizational studies literature to measure several organizational constructs (e.g., Evert, Payne, Moore, & McLeod, 2018; McKenny, Aguinis, Short, & Anglin, 2016; McKenny, Short, & Payne, 2013). Further, research examining reliability in CATA indicates that the algorithm error between CAT Scanner and other common packages such as LIWC and DICTION is generally low and likely attributable to the different features of the package (e.g., whether using phrases or word stems are permitted in dictionaries; McKenny et al., 2016). Because our dictionaries include phrases (e.g., "had a ball", "of service"), CAT Scanner was an appropriate package for our analysis.

Training research indicates that training feedback is multidimensional, suggesting that assessments of satisfaction and utility of training are empirically distinct (Alliger et al., 1997; Brown & Sitzmann, 2011). Guidance regarding validity in CATA suggests that dimensionality be assessed by examining the correlations among the dimensions, with values under 0.50 being indicative of multidimensionality (Short et al., 2010). We found that our measures were significantly correlated with each other (Satisfaction-Utility: $r = -0.03$, $p < 0.05$). However, because none of the correlations approached or exceeded the 0.50 threshold, our measures exhibit the multidimensionality predicted by extant training evaluation research.

To provide an initial concurrent and discriminant validity assessment of our dictionaries, we examined the correlations of our CATA findings with other content analytic measures (Short et al., 2010). We first examined the correlations of our measures with LIWC's discrepancy measure (Pennebaker, Booth, & Francis, 2007). This measure captures words used to note a preference for a state that is different from current reality (e.g., "problem", "should"; Cohn, Mehl, & Pennebaker, 2004). We expect that our satisfaction and utility dictionaries would have significantly weaker relationships with this measure. Our findings matched our expectations: the correlations with satisfaction ($r = 0.01$, n.s.) and utility ($r = 0.04$, $p < 0.05$) were quite small. This provides an indicator of discriminant validity for our satisfaction and utility measures. We also examined the relationship of our measures with LIWC's positive emotion dictionary (Pennebaker et al., 2007). Because expressions of positive emotion in training feedback is likely associated with evaluations of satisfaction, we expect positive correlations between positive emotion, satisfaction, and utility. As expected, positive emotion scores were moderately-to-highly correlated with satisfaction ($r = 0.42$, $p < 0.05$) and utility ($r = 0.18$, $p < 0.05$), and negative emotion scores were weakly correlated with satisfaction ($r = -.03$, $p < .05$) and utility (r

< .001, n.s.). This provides an additional indicator of convergent and discriminant validity for our satisfaction and utility measures.

Table 1

Trainee Reactions Dictionaries for Computer-Aided Text Analysis (CATA)

Construct	Word list
Satisfaction	a real treat, a treat, admiration, admire, admired, admiring, admiringly, adore, adored, adores, adoring, adoringly, agreeability, agreeable, agreeableness, agreeably, amaze, amazed, amazes, amazing, amazingly, amuse, amused, amusedly, amusement, amusements, amuser, amuses, amusing, amusingly, appealed, applaud, applauded, applause, appreciate, appreciated, appreciating, appreciation, appreciations, appreciative, appreciatively, appreciativeness, appreciatory, ate it all up, ate it up, awesome, awesomely, awesomeness, bask, basked, basking, beautiful, beautifully, bed of roses, best, better, blast, blessed, blessing, bliss, blisses, blissful, blissfully, blissfulness, bravissimo, bravo, brighten up, brightened, brightening up, brightens, captivate, captivated, captivates, captivating, captivation, caring, cheer, cheered, cheerful, cheerfully, cheerfulness, cheerier, cheeriest, cheerily, cheeriness, cheering, cheerly, cheers, cheery, comfort, comfortable, comfortableness, comfortably, comforted, comforter, comforting, comfortingly, comforts, commend, commendable, commendably, commendation, commendations, commendatorily, commendatory, commended, commending, commends, compassion, congenial, congeniality, congenially, considerate, contented, contentedly, contentedness, contenting, contently, contentment, cordial, cordially, cordialness, courteous, crowd-pleaser, crowd-pleasing, delectability, delectable, delectably, delectation, delicious, deliciously, deliciousness, delight, delighted, delightedly, delightful, delightfully, delightfulness, delighting, delights, dote on, doted on, dotes on, doting on, dynamic, easygoing, eat it all up, eat it up, eating it all up, eating it up, ecstacy, ecstasies, ecstasy, ecstatic, ecstatically, elate, elated, elatedly, elates, elating, elation, elations, empathetic, empathy, engage, engaged, engages, enjoy, enjoyable, enjoyableness, enjoyably, enjoyed, enjoying, enjoyment, enjoyments, enjoys, enliven, enlivened, enlivening, enlivenment, enlivenments, enlivens, entertain, entertained, entertainer, entertaining, entertainingly, entertainment, entertainments, entertains, enthrall, enthralled, enthralling, enthrallingly, enthrallment, enthrallments, enthralls, enthused, enthusiam, enthusiasm, enthusiastic, excellence, excellent, excellently, exceptional, excited, excitedly, excitement, exciting, exhilarate, exhilarated, exhilarates, exhilarating, exhilaration, exhilarative, fantastic, fantasticality, fantastically, fantasticalness, favorite, favorites, felicitate, felicitated, felicitates, felicitating, felicitation, felicitations, felicities, felicitous, felicitously, felicity, finer, finest, fond, fonder, fondest, fondly, fondness, friendlily, friendliness, friendly, fun, funny, gaiete, gaieties, gaiety, genial, geniality, genially, get a kick out of, getting a kick out of, glad, gladded, gladden, gladdened, gladdening, gladdens, gladder, gladdest, gladding, gladlier, gladliest, gladly, gladness, glee, gleeful, gleefully, gleefulness, glees, gleesome, good, goodly, goodness, got a kick out of, graceful, gracious, grand, grateful, gratification, gratifications, gratified, gratifies, gratify, gratifying, gratifyingly, great, greater, greatest, greatness, greats, groovier, grooviest, groovy, gusto, gustoes, had a

	ball, happier, happiest, happily, happiness, happy, have a ball, having a ball, hit the spot, hits the spot, hitting the spot, humor, humorful, humorous, humorously, humorousness, humorsome, humour, humourous, humours, hunky-dory, incredible, incredibleness, intriguing, intriguingly, is a riot, jollied, jollier, jollies, jolliest, jollification, jollifications, jollified, jollifies, jollify, jollifying, jollily, jolliness, jollities, jollity, jolly, jollying, joy, joyance, joyed, joyful, joyfuller, joyfullest, joyfully, joyfulness, joying, joyous, joyously, joyousness, joys, kind, kindly, kindness, kudos, laugh, likable, likableness, like, likeable, liked, likes, liking, lovable, lovableness, lovably, love, loveable, loveably, loved, lovelier, lovelies, loveliest, lovelily, loveliness, lovely, loves, loving, lovingly, lovingness, marvelous, marvelously, marvelousness, memorable, merrier, merriest, merrily, merriment, merriness, merry, merrymaking, mirth, mirthful, mirthfully, mirthfulness, mirths, neat, nice, nicely, niceness, nicer, nicest, niceties, nicety, perfect, perkier, perkiest, perkily, perkiness, perky, personable, personableness, phenomenal, phenomenally, playful, playfully, playfulness, pleasance, pleasant, pleasanter, pleasantly, pleasantness, pleased, pleases, pleasing, pleasingly, pleasingness, pleasurable, pleasurableness, pleasurably, pleasure, pleasured, pleasureful, pleasuregiving, pleasures, pleasuring, polite, positiveness, positiver, positives, positivest, positivity, praise, praised, propitious, propitiously, propitiousness, rapport, refreshing, refreshingly, regale, regalement, regales, regaling, rejoice, rejoiced, rejoices, rejoicing, relatable, relish, relishable, relished, relishing, revel, reveled, reveling, revelled, revelling, revellings, revelment, revelries, revelry, revels, rocks, satisfaction, satisfactions, satisfactorily, satisfactoriness, satisfactory, satisfied, satisfies, satisfy, satisfying, satisfyingly, savor, savored, savorier, savoriest, savorily, savoriness, savoring, savour, savoured, savourier, savouriest, savouring, savoury, soothing, stimulate, stimulates, stimulating, super, superb, superbly, superbness, superior, superlative, supportive, sweet, sweetly, sweetness, swell, sweller, swellest, terifically, terrific, thoughtful, thoughtfulness, thrill, thrilled, thrillful, thrilling, thrillingly, thrills, upbeat, uplift, uplifted, uplifting, warmhearted, warmness, was a riot, wonderfulness
Utility	adjuvant, adopt, adoptabilities, adoptability, adoptable, adopting, adoption, adoptions, advantage, advantaged, advantageous, advantageously, advantageousness, advantages, advantaging, aid, aidance, aided, aidful, aidfulness, aiding, aids, applicabilities, applicability, applicable, applicably, application, applications, applicative, applicatively, applied, applies, apply, applying, apposite, appositely, appositeness, appreciate, appreciated, appreciation, appropriable, appropriate, appropriately, appropriateness, appropriating, appropriation, appropriations, appropriative, apropos, apt, apter, aptest, aptly, aptness, aptnesses, asset, assets, assist, avail, availed, availing, avails, befit, befits, befitted, befitting, beneficence, beneficent, beneficently, benefices, beneficial, beneficially, beneficialness, beneficient, beneficing, beneficium, benefit, benefited, benefiting, benefits, benefitted, benefitting, conducive, conduciveness, constructive, constructively, constructiveness, effective, effectively, effectiveness, effectual, effectuality, effectually, efficacious, efficaciously, employ, employability, employable, employing,

	enable, enablement, enables, enabling, enhance, enhanced, enhancement, enhancing, enlighten, enlightening, enlightenment, enrich, enriched, enriching, enrichment, equip, equipped, equipping, equips, essential, essentialness, exercisable, expedience, expediency, expedient, expediential, expediently, expedients, feasibility, feasible, feasibleness, feasibly, gain, gained, gainful, gainfully, gainfulness, gaining, gains, germane, germanely, germaneness, handier, handiest, handiness, handy, help, helped, helpful, helpfully, helpfulness, helping, helps, implement, implementation, implemented, implementing, importance, important, improve, improved, improvement, improvements, improving, incorporate, incorporating, incorporation, influencing, influential, influentially, informational, informative, inspiration, inspirational, inspired, inspiring, instrumental, instrumentalities, instrumentality, instrumentally, integrate, integrating, integration, invaluable, life changing, many uses, meritable, meritorious, meritoriously, meritoriousness, motivate, motivated, motivates, motivating, motivation, motivational, multiple uses, multipurpose, of service, pertain, pertained, pertaining, pertains, pertinence, pertinencey, pertinencies, pertinency, pertinent, pertinently, practicabiity, practicabilities, practicability, practicable, practicableness, practicably, practical, practicality, practice, practicing, practising, pragmatic, pragmatical, pragmatically, prepared, profit, profitability, profitable, profitableness, profitably, profited, profiting, profits, purposed, purposeful, purposefully, purposefulness, purposes, purposing, purposive, realism, realistic, realistically, relatable, relevance, relevances, relevancies, relevancy, relevant, relevantly, rewarding, serviceability, serviceable, serviceableness, serviceably, solutions, tangible, timeliness, timely, usability, usable, usableness, usably, usage, usages, use, useability, useable, useably, viabilities, viability, viably, vitally, vitalness

Supplemental Material References

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, *50*(2), 341–358. https://doi.org/10/dn84s7

Brown, K. G., & Sitzmann, T. (2011). Training and employee development for improved performance. In S. Zedeck (Ed.), *APA Handbook of Industrial and Organizational Psychology* (Vol. 2, pp. 469–503). American Psychological Association. https://doi.org/10.1037/12170-016

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693. https://doi.org/10.1111/j.0956-7976.2004.00741.x

Ellis, L. (1994). *Research methods in the social sciences*. WCB Brown & Benchmark.

Evert, R. E., Payne, G. T., Moore, C. B., & McLeod, M. S. (2018). Top management team characteristics and organizational virtue orientation: An empirical examination of IPO firms. *Business Ethics Quarterly*, *28*(4), 427–461. https://doi.org/10.1017/beq.2018.3

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Addison-Wesley.

McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. (2018). What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management*, *44*(7), 2909–2933. https://doi.org/10.1177/0149206316657594

McKenny, A. F., Short, J. C., & Newman, S. M. (2012). CAT Scanner (Version 1.0) [Software]. Available from http://www.catscanner.net/

McKenny, A. F., Short, J. C., & Payne, G. T. (2013). Using computer-aided text analysis to elevate constructs: An illustration using psychological capital. *Organizational Research Methods*, *16*(1), 152–184. https://doi.org/10.1177/1094428112459910

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC 2007*. The University of Texas at Austin. https://www.liwc.net/LIWC2007LanguageManual.pdf

Rodale, J. I. (1978). *The synonym finder*. Rodale.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods*, *13*(2), 320–347. https://doi.org/10.1177/1094428109335949

Zachary, M. A., McKenny, A., Short, J. C., & Payne, G. T. (2011). Family business and market orientation: construct validation and comparative analysis. *Family Business Review*, *24*(3), 233–251. https://doi.org/10.1177/0894486510396871