Analyzing Qualitative Trainee Reactions:
A Phenomenological Study of Instructional Designer Practices

Stephen J. Naso
B.S.B.A., Emmanuel College, 2009
M.Ed., University of Massachusetts Boston, 2013

Submitted to the Graduate Department and Faculty of the School of Education of
Baker University in partial fulfillment of the requirements for the degree of
Doctor of Education in Instructional Design and Performance Technology

Wendy Ann Gentry
Wendy A. Gentry, Ph.D.
Major Advisor

Regena M. Aye
Regena M. Aye, Ed.D.

Mike Gentry
Mike Gentry, Ed.D.

Date Defended: November 29, 2021

Abstract

It is well established that Level 1 of the Kirkpatrick model is the most used training evaluation method and is typically carried out by distributing reactionnaires (i.e., smile sheets) at the end of a training program. While the open-ended questions commonly included in reactionnaires can help capture critical program evaluation information, there is a lack of guidance in the literature for instructional designers on analyzing this type of data, referred to as qualitative reactions. Thus, there were two purposes for this qualitative study: first, to examine the methods used by instructional designers to analyze qualitative reactions, and second, to examine the underlying reasons for using those methods of analysis. Semi-structured interviews were conducted with 13 instructional designers. An analysis of the interview data revealed three themes regarding the methods used to analyze qualitative reactions: most participants used at least one surveying platform to collect the qualitative reactions; most participants used a range of tactics to look for keywords, patterns, or themes in the data; and most participants reviewed the data as a team. Three themes also emerged regarding the underlying reasons for the methods of analysis: all participants had at least one primary motive for analyzing the data; most participants incorporated their unique skills, knowledge, and abilities when analyzing their data sets; and most participants had a support system in place to help with the data analysis. The results of this study provide insight into an under-researched aspect of Level 1 evaluation and can be used to help educate instructional designers on qualitative data analysis (QDA) practices.

Acknowledgments

The completion of my dissertation would not have been possible without the support of many individuals, and while I am appreciative of all of them, special thanks go to several individuals. I would like to express my deepest gratitude to Dr. Gentry, my major advisor, and Dr. Cho, my research analyst. Their help cannot be overestimated. Many thanks to my committee, who provided their time, encouragement, and expertise. I am also grateful to Bridget, whose profound belief in my abilities so many years ago set the stage for my career in this field that I love so much. Thanks must also go to my grandmother for being my biggest cheerleader and Kevin for his helpful advice and patience throughout this project.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Since its introduction more than 60 years ago, the Kirkpatrick model has persisted as a popular training program evaluation model for instructional designers. Throughout the model's lifespan, an extensive body of literature has emerged, with countless researchers having conducted studies that incorporate the model as a critical framework or having conducted studies on the model itself. The current study falls under the latter category and explores an under-researched aspect of the model. This chapter introduces the study and includes background information, the statement of the problem, the study's purpose and significance, the study's delimitations and assumptions, the study's research questions, and definitions of terms that occur throughout the study.

Background

Beginning in November 1959, Donald L. Kirkpatrick published a series of articles in the *Journal of the American Society of Training Directors* that outlined steps for evaluating training programs (Kirkpatrick & Kirkpatrick, 2009). He postulated that one performs his prescribed evaluation techniques in sequential order to ultimately collect data that presents a holistic and accurate measurement of the effectiveness of a training program. The first step was to measure the participants' overall reaction to the training. The second step was to measure the increase in participants' knowledge as a result of the training. The third step was to measure participants' on-the-job changes in behavior as a result of the training. The fourth step was to measure the training's impact on the business. While Donald L. Kirkpatrick initially referred to his four techniques as *steps*, instructional designers began referring to the steps as *levels*, and the levels ultimately

came to be known as the *Kirkpatrick model* (Kirkpatrick & Kirkpatrick, 2009), as displayed in Figure 1.

Level 1: Reaction	To what degree participants react favorably to the learning event.
Level 2: Learning	To what degree participants acquire the intended knowledge, skills, and attitudes based on their participation in the learning event.
Level 3: Behavior	To what degree participants apply what they learned during training when they are back on the job.
Level 4: Results	To what degree targeted outcomes occur as a result of the learning event(s) and subsequent reinforcement.

Figure 1. The Four Levels of the Kirkpatrick Model. From *Kirkpatrick Then and Now: A Strong Foundation for the Future* (p. 3) by J. D. Kirkpatrick and W. K. Kirkpatrick, 2009, Saint Louis, MO: Kirkpatrick Partners, LLC. Copyright 2009 by James D. Kirkpatrick and Wendy Kayser Kirkpatrick.

Although the Kirkpatrick model is the most popular training program evaluation model (Bates, 2004; Brinkerhoff & Dressler, 2015; Holton, 1996; Kaufman & Keller, 1994; Pulichino, 2007), instructional designers do not always incorporate every level of the model into their program evaluation strategies. Because of the ease with which it can be implemented, Level 1 evaluation is the most commonly used training evaluation method (Association for Talent Development, 2016; Brown, 2005; Chapman, 2000; Gomez, 2003; Hypes, 2016; Moller & Mallin; 1996; Phillips, 2000), and in some cases, the only form of evaluation used to make program decisions (Morgan & Casper, 2000). Instructional designers routinely conduct Level 1 evaluation using questionnaires that

collect trainee reactions to learning events, which are commonly referred to as *smile sheets*, *happy sheets*, *reaction sheets*, or *reactionnaires* (Lee, 1998; Newby, 1992; Thalheimer, 2016b), with the latter being the term used throughout this paper. Reactionnaires customarily include a varied mix of closed- and open-ended questions that are dependent on the program being measured and the purpose of the evaluation (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Lee & Pershing, 2002; Phillips & Phillips, 2016).

Statement of the Problem

Although it is well-established that instructional designers conduct training program evaluations for a broad range of reasons (see Brown, 2005; Clegg, 1987; Kirkpatrick & Kirkpatrick, 2015b; Kraiger, 2002; Moller & Mallin, 1996; Newstrom, 1978; Phillips & Phillips, 2010; Pulichino, 2007; Russ-Eft & Preskill, 2009), that the Kirkpatrick model is the most widely used training program evaluation model, and that organizations that evaluate training programs generally restrict their evaluation practices to Level 1, there is a relative lack of research on Level 1 evaluation itself. Furthermore, while some of the most crucial training program evaluation information is captured via reactionnaire questions that yield qualitative data (J.L. Pershing, 2006; Kirkpatrick, 2008; Kirkpatrick & Kirkpatrick, 2009; Thalheimer, 2016c)—such as questions asking about the value of the program or how it could be improved—no literature or research appears to exist on how instructional designers analyze qualitative data derived from reactionnaires. A broad range of scholarly and non-scholarly works detail the value of the data and methods for obtaining the data, such as how to structure questions and when

to deliver surveys, but all stop short in discussing the ways to analyze the qualitative results.

Purpose of the Study

Given the aforementioned gap in research, there were two primary purposes for this phenomenological study. The first purpose was to examine the methods used by instructional designers to analyze qualitative data obtained via reactionnaires, which are referred to as *qualitative reactions* (Harman, Ellington, Surface, & Thompson, 2015) throughout this paper. The second purpose was to examine the underlying reasons for the use of the methods of analysis.

Significance of the Study

This study adds to our understanding of how instructional designers analyze qualitative reactions. This study may be significant in four ways. First, it may be helpful to a broad array of instructional designers who need to analyze training reaction data to determine learner satisfaction, improve training programs, or justify the value or worth of training programs. Second, it can provide more direction to instructional designers whose evaluation work is guided by core standards and competencies, such as the standards and competencies upheld by the International Board of Standards for Training, Performance, and Instruction (IBSTPI). IBSTPI maintains a set of standards for instructional designers that includes a specification that they "evaluate instructional and non-instructional interventions" and "revise instructional and non-instructional solutions based on data" (IBSTPI, 2013, p. 6). Third, it can assist instructional designers and organizations that are accredited by professional instructional design associations and expected to uphold the association's evaluation standards. For example, organizations

that are accredited by the International Association for Continuing Education and Training (IACET) are expected to:

> Have a process for analyzing learning event evaluation results and sharing them with instructors, instructional design and development staff, program evaluators and administrators to ensure these results are incorporated into continuous process improvement for the specific learning event and future learning events. (IACET, 2017, p. 20)

Fourth, it may be useful to professional instructional design associations such as IBSTPI and IACET in refining their training program evaluation standards and competencies.

Delimitations

This study recognized five delimitations. First, a qualitative research approach was selected because of the lack of existing research on the topic. The lack of existing research on the topic would have created difficulty in developing valid and reliable quantitative data collection instruments. Second, the sample for this study was limited to instructional designers who evaluated at least two training programs using reactionnaires that included at least one qualitative question. Third, the study investigated each participant's experience with the phenomenon over the 6 months that preceded the interview. This period was selected because many instructional designers were changing how they delivered training toward the beginning of the COVID-19 pandemic. It was believed that 6 months before each interview was a reasonable point at which each participant had become established with any new or revised training standards or practices. Fourth, the study only investigated instructional designers' analysis of qualitative reactions and did not investigate their analysis of quantitative data or any

connections between the qualitative and quantitative data. Fifth, the study did not explore how participants used the data they analyzed.

Assumptions

This study included two assumptions. The first assumption was that the participants understood the interview questions. The second assumption was that the participants provided answers to interview questions as accurately as possible, given their recall of the subject matter.

Research Questions

The following research questions (RQ) guided this study:

RQ1: How do instructional designers analyze qualitative reactions?

RQ2: What rationale do instructional designers provide for the methods they use to analyze qualitative reactions?

Guided by pragmatism, the researcher viewed the research problem from the lens of a human performance improvement practitioner and selected a qualitative research approach. This approach utilized a phenomenological methodology where interviews with the participants were used to help answer the research questions.

Definition of Terms

This study used the following definitions of key terms and concepts:

Evaluation: Definitions of evaluation abound and can be generic and straightforward, such as "determining whether objectives have been achieved" (Stufflebeam & Coryn, 2014, p. 6), or more specific and elaborate:

> Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the

> program, improve program effectiveness, and/or inform decisions about future programming. (Patton, 2015, p. 18).

Throughout this paper, evaluation refers to the overarching act of measuring the effectiveness and impact of training programs.

Evaluation versus research: Evaluation and research are not synonymous; they are separate and distinct practices (King, Stevahn, Ghere, & Minnema, 2001). For this paper, evaluation is considered a form of applied research based on Chen's (2018) positioning of the difference between the two:

> The fundamental difference between evaluation and research lies in the purpose and motivation behind the work. The goal behind research is to contribute to a body of knowledge and theory in a field.... In contrast, the focus of evaluation, as a form of applied research, is to judge merit or quality as determined by the interests of various stakeholders. (p. 630)

Given this position, *evaluation* is primarily used throughout this paper in the context of training program evaluation, while *research* is primarily used in the context of academic research.

Instructional design and instructional designers: The practice of instructional design occurs in many organizations, including those in business and industry, education, government or military, healthcare, and other entities (Klein & Kelly, 2018). Within these organizations, instructional design practitioners often hold a variety of titles. For example, in a study that sought to uncover how training professionals implement instructional technology and the competencies needed to do so, Furst-Bowe (1996) analyzed 147 completed questionnaires from members of the International Society for

Performance Improvement and found that the respondents held 40 different job titles. Nevertheless, although instructional designers hold various titles, they share common underlying goals, practices, and skillsets. Thus, this study adopts Reiser's (2018) definition of the field and its practitioners, which is both broad and specific enough for this study:

> The field of instructional design and technology (also known as instructional technology) encompasses the analysis of learning and performance problems, and the design, development, implementation, evaluation and management of instructional and non-instructional processes and resources intended to improve learning and performance in a variety of settings, particularly educational institutions and the workplace. Professionals in the field [of] instructional design and technology often use systematic instructional design procedures and employ instructional media to accomplish their goals. Moreover, in recent years, they have paid increasing attention to non-instructional solutions to some performance problems. Research and theory related to each of the aforementioned areas is also an important part of the field. (pp. 4–5)

Level 1 evaluation: Level 1 evaluation refers to a single level of evaluation within the larger Kirkpatrick model that is focused on investigating trainee reactions to a training program. While reactionnaires are the most commonly used technique to measure reactions, other techniques may be used at this level, such as collecting verbal feedback. Therefore, *Level 1 evaluation* is used when discussing the general act of evaluating trainee reactions, while *reactionnaire* is used when discussing the specific means by which evaluators do so.

Training program: A training program is any formal learning event. A formal learning event is goal-oriented, planned, and structured.

Qualitative data analysis (QDA): QDA refers to "the process of closely scrutinizing and interpreting qualitative data with the aim of transforming it into findings and conclusions" (Hart & Achterman, 2017, p. 1).

Organization of the Study

The current study is divided into five chapters. Chapter 1 introduced the study and included background information, the statement of the problem, the study's purpose and significance, the study's delimitations and assumptions, the study's research questions, and definitions of terms that occur throughout the study. Chapter 2 examines the literature that is relevant to the study and includes the conceptual and theoretical frameworks that underpin the study. Chapter 3 details the methods used to investigate the phenomenon, including the overall research design, the setting and sampling procedures, the instruments and data collection procedures, the data analysis and synthesis processes, the measures used to guarantee the study's reliability and trustworthiness, the researcher's overall role in the study, and the study's limitations. Chapter 4 presents the study's results, including the themes that emerged via the methods outlined in Chapter 3. Chapter 5 provides a summary of the study; the findings related to the literature; and overall conclusions, including implications for action, recommendations for future research, and concluding remarks.

Chapter 2

Review of the Literature

This chapter is divided into four primary sections that present a review of the literature that is relevant to this study. The first section positions training evaluation as a diverse activity for instructional designers and explains why and when these practitioners evaluate training. The second section explores the research conducted on the Kirkpatrick model and highlights the model's popularity with instructional designers, who, at a minimum, customarily conduct Level 1 evaluations via reactionnaires. The third section recaps Kirkpatrick's original guidelines for measuring trainee reactions—which have remained mostly unchanged since they were first introduced more than 60 years ago—and provides an overview of the design of reactionnaires, which are dependent on the program being measured but generally include a mix of closed- and open-ended questions. This section also establishes a connection between the lack of guidance for instructional designers in the literature on analyzing qualitative reactions and the supposition that we can draw from the QDA practices established by researchers. Finally, the fourth section introduces the conceptual and theoretical frameworks used in this study to investigate how instructional designers analyze qualitative reactions.

Training Evaluation as a Diverse Activity

The practice of evaluating training involves "a planned effort to measure what happens in training, how it affects trainee knowledge, skills, abilities, and performances, and training's impact on organizational outcomes" (Gomez, 2003, p. 23). Although this statement provides a straightforward description of training evaluation and highlights the primary purpose of evaluation—to determine the effectiveness of training— instructional

designers conduct training evaluation for several ancillary reasons that are unique to and driven by the underlying goals and objectives of the instructional designers, their departments, and their organizations. For example, and in many instances, instructional designers evaluate training so they can gather data for decision-making purposes (Brown, 2005; Clegg, 1987; Kirkpatrick & Kirkpatrick, 2015b; Newstrom, 1978; Russ-Eft & Preskill, 2009), with the most popular decisions revolving around program improvement and program continuation (Moller & Mallin, 1996). In other instances, instructional designers evaluate training to demonstrate a training department's value to and impact on the larger organization (Clegg, 1987; Kirkpatrick & Kirkpatrick, 2006; Kirkpatrick & Kirkpatrick, 2015b), which may include measuring the return on investment (ROI) for training initiatives and monetarily proving the department's worth (Clegg, 1987; Kirkpatrick & Kirkpatrick, 2006; Phillips & Phillips, 2010; Pulichino, 2007). In even other instances, instructional designers evaluate training to capture valuable data that can be used to help market their training programs (Brown, 2005; Kraiger, 2002). These reasons, among others, drive instructional designers to evaluate training at one or more points over the lifespan of a training initiative. The various points at which training evaluation might occur can be funneled into three specific timeframes that are inexplicably linked to three distinct types of evaluation as defined by DeVaughn and Stefaniak (2020): *formative*, *summative*, and *confirmative*.

Scriven (1966) originally distinguished between formative and summative evaluation, which are standard terms found today in the literature on training evaluation. Formative evaluation occurs while a training program is in development, while summative evaluation occurs after the program has been delivered. Misanchuk (1978),

having found that instructional designers had started to obfuscate the formative and summative terminology, and having considered the individuality of instructional design and evaluation projects, introduced confirmative evaluation as a third type of evaluation. The major distinction between summative and confirmative evaluation is time (Hellebrandt & Russell, 1993); whereas summative evaluation occurs just after the training program has been delivered, confirmative evaluation occurs after the training program "has been put into practice for a period of time and is now up for review" (Misanchuk, 1978, p. 16).

While the range of reasons for conducting training evaluation and the points at which evaluation occurs vary, it is evident that the practice has become an essential aspect of most instructional design models (Eseryel, 2002). Phillips and Phillips (2016) identified nearly 30 training evaluation models that instructional designers can draw from. Although the menu of evaluation models is plentiful for instructional designers, none is more popular than the Kirkpatrick training evaluation model, which instructional designers have embraced and adopted as the standard approach for evaluating their programs (Bates, 2004; Brinkerhoff & Dressler, 2015; Holton, 1996; Kaufman & Keller, 1994; Pulichino, 2007).

The Kirkpatrick Model

A review of the literature reveals that the Kirkpatrick model is the most widely used training evaluation model by instructional designers and that lower levels of the model (i.e., Levels 1 and 2) are used more frequently than the higher levels of the model (i.e., Levels 3 and 4) (see Blanchard, Thacker, & Way, 2000; Foreman, 2008; Kennedy, Chyung, Winiecki, & Brinkerhoff, 2004; Long, 1990; Pulichino, 2007). From its initial

introduction to instructional designers through today, the Kirkpatrick model has remained popular due to its apparent simplicity (Alliger & Janak, 1989; Alliger et al., 1997; Bates, 2004; Chapman, 2000; Giangreco, Carugati, & Sebastiano, 2010; Newstrom, 1978). However, researchers equally perceive this simplicity as one of the model's primary shortcomings, along with being incomplete and lacking causation and correlation among the levels (Alliger & Janak, 1989; Bates, 2004; Holton, 1996; Guerra-López, 2008; McLinden & Boone, 2009). These shortcomings have prompted researchers to question the model's merit and soundness. Moreover, the lack of causation and correlation among levels has become the model's primary deficiency, inducing researchers to investigate the subject. However, the literature reveals contradictory conclusions, as evidenced by the following:

- Clement (1982) conducted a study to ascertain whether reactions could predict learning and behavior outcomes and found that reactions were positively related to learning, and learning was positively related to behavior.
- Alliger and Janak (1989) conducted a meta-analysis to determine the validity of three assumptions: that each succeeding level provided increasingly informative data, that each level caused the next, and that there were positive correlations among the levels. Their findings did not provide adequate support for the assumptions.
- Building on the work of Alliger and Janak (1989), Alliger et al. (1997) conducted a meta-analysis that explored the relationships among a set of training evaluation criteria and the first three levels of the model. Their findings provided only modest support for correlations among the levels.

- Warr, Allan, and Birdi (1999) examined the relationships among the first three levels of the model, including the individual and organizational predictors of each level. The authors discovered a significant relationship between reactions and learning but an insignificant relationship between reactions and behavior.
- In an investigation of trainee reactions as predictors of learning, Tan, Hall, and Boyce (2003) found that negative reactions predicted learning outcomes.
- Advancing Brown's (2005) work on the nomological network of reactions, Sitzmann, Brown, Casper, Ely, and Zimmerman (2008) conducted a meta-analysis to determine the degrees to which a set of antecedents predicted reactions and the degrees to which reactions were related to both learning and organizational outcomes. Their findings supported the link between reactions and learning.
- Blume, Ford, Baldwin, and Huang (2010) updated the meta-analysis conducted by Alliger et al. (1997). Their findings were consistent with Alliger et al.'s in that reactions did not significantly predict learning.
- Kim, Park, Lavelle, Kim, and Chaudhuri (2020) explored the relationships among trainee antecedents, trainee reactions, and training outcomes. Their findings supported those of Sitzmann et al. (2008).

The attention paid to the Kirkpatrick model's criticisms has led some researchers to develop additional evaluation models that complement or expand on the Kirkpatrick model (Russ-Eft & Preskill, 2009; Dessinger & Moseley, 2006; Phillips & Phillips, 2016). To illustrate, Kaufman and Keller (1994) revised Kirkpatrick's Level 1 to include the evaluation of training inputs and added a fifth level to measure societal contributions,

while Phillips (Phillips & Phillips, 2016) added a fifth level to the model to measure training ROI. Other researchers stand by the original model or point to the need for entirely new evaluation methods, such as Giangreco et al. (2010), who argued that the current state of our economy is much different than when the model was first introduced in an industrial era; thus, the model's criticisms are irrelevant today, and instead of modifying the model to meet our evolved training and development needs, we must develop new evaluation tools. Giangreco et al. also presented rebuttals to the core limitations of the Kirkpatrick model. On the criticism that the model is oversimplified, the authors countered that the model is more complex in our post-industrial era because of the resources it requires, such as the amount of time needed to evaluate all four levels, and that today's organizations may not be equipped with those resources. On the model being incomplete, the authors agreed but acknowledged that addressing this issue would only lead to more complexity, thereby limiting its use. On the lack of proof of causation among the levels, Giangreco et al. acknowledged this absence but indicated that causality may not be necessary today, for it is more important to evaluate according to the program's goals.

Research on Level 1 evaluation. Despite the widespread use of the Kirkpatrick model, and even though instructional designers more frequently conduct Level 1 and Level 2 evaluations, a relatively limited body of research exists on Level 1 evaluation itself. As suggested by Giangreco, Sebastiano, and Peccei (2009), the lack of research in this area may result from researchers viewing the higher levels of evaluation as more important in assessing program effectiveness, thus reducing the importance of Level 1 evaluation and deeming research unworthy. While the research that does exist on Level 1

evaluation is focused mainly on exploring the correlation between trainee reactions and learning outcomes, the research that falls outside of this scope can be segmented into two groupings. The first grouping of non-correlation-related research on Level 1 evaluation surrounds the design and development of the evaluation itself. Namely, researchers such as Lee (1998), Lee and Pershing (2002), and Pershing and Pershing (2001) have conducted studies to determine the optimal design criteria of reactionnaires; their findings are incorporated into a latter section of this literature review covering the design of reactionnaires. The second grouping of non-correlation-related research on Level 1 evaluation surrounds the nomological underpinning of the trainee reactions themselves. It must be noted that in this area, some researchers have grouped this nomological research with outcomes research, such as Sitzmann et al. (2008) and Kim et al. (2020), who have investigated the relationship between antecedents and reactions, and reactions and training outcomes. However, other researchers such as Giangreco et al. (2009) have focused solely on investigating the antecedents, underlying constructs, and underlying factors that impact reactions. Giangreco et al. (2009) examined the antecedents of trainee reactions by surveying 2,697 participants from a wide range of organizations that participated in a training program offered by a prominent Italian training agency. Giangreco et al. found that trainee reactions were driven by trainees' perceptions of the organization of the training, perceptions of the instructor's performance, and perceptions of the usefulness of the training. In another example, Morgan and Casper (2000) studied the underlying constructs and factors that impact reactions by collecting and analyzing participant reactions from 9,128 government agency employees. Morgan and Casper found that reactions were tied to six constructs: satisfaction with the instructor, overall

satisfaction with the training, satisfaction with testing, utility of training, satisfaction with materials, and satisfaction with the course structure. In a third example, Glerum, Joseph, McKenny, and Fritzsche (2020) analyzed over 10,000 trainee reactions that they obtained from a professional development company that offered graduate-level training courses to educators. Glerum et al.'s findings suggested that course trainers had a more significant influence on trainee reactions than course content. As a final example of research conducted on trainee reactions—and in what appears to be the only study investigating trainee responses to open-ended questions—Harman et al. (2015) conducted three field studies that involved hundreds of members of a large military organization who participated in dozens of classes at the organization. The major findings in that study indicated that specific individual and situational factors impacted the likelihood that participants would respond to open-ended questions.

Popularity of Level 1 evaluation. Although a review of the literature reveals mixed findings on the Kirkpatrick model's soundness and an overall lack of research on Level 1 evaluation itself, the use of Level 1 evaluation—which is generally conducted via reactionnaires—has remained prevalent across many industries over the past three decades. For example, Moller and Mallin (1996) surveyed 191 instructional designers on their use of the Kirkpatrick levels and found that 89.5% of the designers evaluated at Level 1; these instructional designers worked in industries including academia, consulting, government, healthcare, military, retail, transportation, and utility/communications. Phillips (2000) surveyed five organizations—an automotive manufacturer, a banking company, a furniture manufacturing company, a health care company, and a property and casualty insurance company—on their evaluative practices

and found that, on average, 84% of the training programs these organizations offered were evaluated at Level 1. Gomez (2003) surveyed 52 financial services organizations and found that 87.29% of training programs were evaluated at Level 1. The Association for Talent Development (2016) categorized the evaluation efforts of 199 participating organizations across a wide array of industries and found that in 2009, 92% of organizations evaluated at Level 1, while in 2015, that number slightly dropped to 88%. In a systematic review of literature on the federal government's training evaluation practices, Hypes (2016) found that most Federal agencies evaluated their training programs at Level 1. Thus, these studies' findings indicate that Level 1 evaluation is the predominant training program evaluation practice for the training industry as a whole.

Value of Level 1 evaluation. Despite the limited research on Level 1 evaluation and the generally accepted supposition that higher levels of evaluation yield more useful data (Alliger & Janak, 1989), instructional designers' widespread use of Level 1 evaluation signifies there is value in conducting this type of evaluation. Level 1 evaluation is useful for three primary reasons. First, reactions help determine customer (i.e., learner) satisfaction (Brinkerhoff, as cited in Thalheimer, 2016a; Thalheimer, 2016b; Kirkpatrick, 2006; Phillips & Phillips, 2010; Phillips & Phillips, 2016). Kirkpatrick (2006) pointed out that some organizations do not consider reaction to be important and therefore do not measure it; however, he cautions against this and reinforces its importance as at least a measure of customer satisfaction. To emphasize this point, he recalled that the University of Wisconsin Management Institute offered a 10-day program where the first 5 consecutive days were followed by a 6 week break and then another 5 consecutive days. Kirkpatrick explained that because participants were

satisfied with the first part of training, they returned for the second. Brinkerhoff (as cited in Thalheimer, 2016a) echoed Kirkpatrick on the usefulness of reactions for customer satisfaction and added that it offers a means by which participants can provide negative feedback that might otherwise be shared directly with more senior individuals in an organization, which could have more damaging effects. Phillips and Phillips (2010) positioned reaction as a critical measure of customer satisfaction that is directly tied to the second form of value: that reactions can inform the design and improvement of training programs (Brinkerhoff, as cited in Thalheimer, 2016a; Thalheimer, 2016b; Kirkpatrick & Kirkpatrick, 2015b; Kraiger, 2002; Morgan & Casper, 2000; Pershing & Pershing, 2001; Phillips & Phillips, 2010; Phillips & Phillips, 2016). For instance, in the case of a pilot program, an instructional designer might solicit input on recommended changes before a broader rollout, or given an existing program, an instructional designer might seek feedback on the program's content, relevance, or facilitators, among other things. Third, reactions are useful in substantiating the justification, value, or worth of a program—or even the justification, value, or worth of an overall training organization (Kirkpatrick & Kirkpatrick, 2015b; Morgan & Casper, 2000; Phillips & Phillips, 2010; Phillips & Phillips, 2016; Thalheimer, 2016b). "Like any other department in an organization, training is not exempt from showing how the resources allocated to them have been put to use" (Kirkpatrick & Kirkpatrick, 2015b, p. 1). Therefore, training organizations can use trainee reactions to market their successes to intended audiences (Kraiger, 2002). Moreover, the success of some training organizations is based solely or in part on trainee reactions (Phillips & Phillips, 2010), which management might use to

make decisions about that training organization or its programs or both (Kirkpatrick & Kirkpatrick, 2009).

Reactionnaires

Standards for measuring reaction. The obtainment of trainee feedback precedes Donald L. Kirkpatrick's work on measuring trainee reaction (Kirkpatrick & Kirkpatrick, 2009; McLinden & Boone, 2009). According to Donald L. Kirkpatrick's account of the origins of the methodology, training *comment sheets* were already in use in the 1950s for programs delivered through the Management Institute at the University of Wisconsin (Kirkpatrick & Kirkpatrick, 2009). In fact, in his first article published in the *Journal of the American Society of Training Directors* in November 1959, he noted that most industry training directors already measured reaction because of the ease in so doing. This practice remains true today, as the primary instrument used to measure reactions are end-of-course surveys (i.e., reactionnaires) (Brown, 2005; Guerra-López, 2008; Lee, 1998; McLinden & Boone, 2009).

Although reactionnaires had become a popular evaluation instrument by 1959, most instructional designers failed to adhere to any standards (Kirkpatrick & Kirkpatrick, 2009). Thus, Donald L. Kirkpatrick established five core guidelines for measuring reaction that have remained relatively unchanged since his articles were first published (Kirkpatrick, 2006; Pulichino, 2007):

1. Determine what you want to find out.
2. Use a written comment sheet covering those items determined in step one above.
3. Design the form so that the reactions can be tabulated and quantified.

4. Obtain honest reactions by making the forms anonymous.
5. Allow the conferees to write in additional comments not covered by the questions that were designed to be tabulated and quantified. (Kirkpatrick & Kirkpatrick, 2009, p. 40)

Question types and frequencies. The questions included in reactionnaires are primarily dependent on the program being measured—as opposed to being generic to a range of programs—and the purpose of the evaluation (Alliger et al., 1997; Lee & Pershing, 2002; Phillips & Phillips, 2016). In preparation for a webinar discussing ways to build better reactionnaires, Kirkpatrick and Kirkpatrick (2015a) surveyed more than 300 readers of a newsletter distributed by their company, Kirkpatrick Partners, and found that more than half of respondents customized reactionnaire questions based on the program being evaluated. Additionally, while facilitating the webinar, the presenters conducted a poll to determine how many questions participants included in their reaction sheets. Of the 340 webinar participants, 13.8% (n = 47) responded that they included one to five questions, 30% (n = 102) included six to ten questions, 12.7% (n = 43) included eleven or more questions, and 3% (n = 10) indicated that it varies depending on the program or content; the remaining respondents either did not use reactionnaires or chose not to respond to the poll. These findings were consistent with those of Lee and Pershing (2002), who determined that reactionnaires generally included 6–15 questions.

Reactionnaire questions include a mix of open-ended (i.e., qualitative) and closed-ended (i.e., quantitative) questions, with the most common being the latter (Lee & Pershing, 2002; Pershing & Pershing, 2001; Thalheimer, 2016a). Qualitative questions tend to include short-answer forms, while quantitative questions tend to include Likert

scale questions. Donald L. Kirkpatrick's standard to "design the form so that the reactions can be tabulated and quantified" (Kirkpatrick & Kirkpatrick, 2009, p. 40) paved the way for this customary practice of using Likert scale questions, for the responses to these questions can be easily converted to numbers. Instructional designers regularly use the following Likert response scales as outlined by Pershing and Pershing (2001):

- Agreement (i.e., strongly disagree to strongly agree).
- Frequency (i.e., never to always).
- Importance (i.e., unimportant to very important).
- Satisfaction (i.e., very dissatisfied to very satisfied).
- Effectiveness (i.e., very ineffective to effective).
- Quality expectation (i.e., very poor to very good, or much worse than expected to much better than expected).
- Extent/likelihood (i.e., to no extent, not at all to a very great extent).

Question dimensions. Researchers have attempted to categorize and classify the broad range of reactionnaire questions. For instance, Alliger et al. (1997), building on the work of Warr and Bunce (1995), divided questions by affective and utility measures. Affective measures are related to the enjoyment of training (e.g., "I found this training to be enjoyable"), and utility measures are related to the usefulness of training (e.g., "I found this training to be relevant to my job"). Warr and Bunce (1995) suggested training difficulty as a third type of reaction measure, but in their research found that questions related to this measure were not commonly asked. Sitzmann et al. (2008) included measures related to trainee self-efficacy (i.e., that trainees are confident they can use what they learned in training on the job) and motivation (i.e., that trainees will actually use

what they learned in training on the job). Phillips and Phillips (2016) outlined more than one dozen types of reaction data:

- Participant demographics.
- Logistics and service.
- Readiness.
- Objectives.
- Learning materials.
- Facilitator.
- Media/delivery.
- Value of content.
- Practice and labs.
- Value of problem.
- Planned use.
- Future needs.
- Marketing and registration.
- Open comments.

In the most comprehensive review of reactionnaire questions, Lee and Pershing (2002) found the broad range of reactionnaire questions could be divided into eleven dimensions:

- Program objectives/content.
- Program materials.
- Delivery methods and technologies.
- Instructors/facilitators.

- Instructional activities.
- Program time/length.
- Training environment.
- Planned actions/transfer expectations.
- Logistics/administration.
- Overall evaluation.
- Recommendations for program improvement.

Of these eleven dimensions, Lee and Pershing found that reactionnaires typically included 5–7 dimensions, with the most common being recommendations for program improvement, overall evaluation, planned actions/transfer expectations, instructors/facilitators, and program objectives/content.

"New World" reactionnaires. Although reactionnaires have been in use for decades and have been modified over time, most have generally been trainer-centered rather than learner-centered (Kirkpatrick, 2008). In his attempt to position a "new age" of Level 1 evaluation, James D. Kirkpatrick encouraged instructional designers to do two things: first, shift trainer-centered Likert scale statements such as *the program objectives were clearly defined* to learner-centered statements like *I understood the learning objectives*, and second, add questions to measure engagement and relevance. On this latter suggestion, Kirkpatrick and Kirkpatrick (2015a) affirmed that instructional designers should use Level 1 evaluation to ask questions related to the other levels. The authors suggested instructional designers include a variety of questions, including open-ended survey questions.

Reactionnaire data analysis. When analyzing quantitative trainee reactions, instructional designers customarily convert the Likert scale responses to numbers (i.e., scores) and then use those numbers to calculate averages (Thalheimer, 2016a). However, some researchers have elected to use alternative quantitative data analysis methods. In particular, McLinden and Boone (2009) suggested approaching this practice in a more psychometrically sound manner by conducting a Rasch analysis. The authors believed this method was a better alternative to simply calculating averages because it "takes into account the fact that raters differ in their use of the rating scale: some people are easy raters and tend to use the high end of the scale, and some are hard raters and tend to use the low end of the scale" (McLinden & Boone, 2009, p. 11). When analyzing qualitative trainee reactions, it appears most instructional designers choose to present the data in narrative form. To illustrate, in a study that evaluated an online faculty development program at Boise State University, Chen, Lowenthal, Bauer, Heaps, and Nielsen (2017) used several reactionnaires that included a mix of quantitative and qualitative questions. The researchers reported the averages for the quantitative data and narratives for the qualitative reactions.

Despite the usefulness of qualitative questions, little attention has been paid to the analysis of this data. Donald L. Kirkpatrick's original guidance for Level 1 evaluation favored quantitative questions over qualitative questions, but he indicated that reactionnaires should include an open-ended question that asks trainees what would have improved the program (Kirkpatrick & Kirkpatrick, 2009). In his positioning of "New World" reactionnaires, James D. Kirkpatrick amplified this original recommendation and

suggested instructional designers include more open-ended questions to capture more useful data:

> Keep the ever-popular questions regarding what the participants liked and didn't like, but consider adding some or all of the following to really get the point across:
>
> - What were the three most important things you learned from this session?
> - From what you learned, what do you plan to apply back at your job?
> - What kind of help might you need to apply what you learned?
> - What barriers do you anticipate you might encounter as you attempt to put these new skills into practice?
> - What ideas do you have for overcoming the barriers you mentioned?
> - What ultimate impact do you think you might contribute to the organization as you successfully apply what you learned? (Kirkpatrick, 2008, p. 5)

Although prominent evaluation researchers indicate the usefulness and value of open-ended questions (viz., J.L Pershing, 2006; Kirkpatrick, 2008; Kirkpatrick & Kirkpatrick, 2009; Thalheimer, 2016c), they fail to outline methods for analyzing this data. In Lee and Pershing's (2002) research on the design of reactionnaires, the authors found that most reactionnaires fail to adhere to five basic design criteria: introductions, directions, and closing statement; response format; question construction; layout; and data analysis. Although their research revealed a range of open-ended questions used by instructional designers, and although the authors offered recommendations for the first four criteria, Lee and Pershing stopped short of exploring the data analysis criterion.

Kirkpatrick and Kirkpatrick (2009) noted that "it becomes very difficult to summarize comments and to determine patterns of reaction" (p. 43) and only indicated that "an executive report might include an aggregate of participant satisfaction scores and a few representative testimonials" (Kirkpatrick & Kirkpatrick, 2015b, p. 10). Some books dedicated to evaluation even devote entire chapters to methods for analyzing quantitative data (e.g., Phillips & Phillips, 2016; Thomas, 2006) but only provide simple guidelines for analyzing qualitative data (e.g., Lee, 2006), or avoid discussing open-ended questions altogether (e.g., Pershing & Pershing, 2001). Newby (1992) appears to provide the most thorough guidance on analyzing qualitative reactions in a single paragraph that directs instructional designers to review each response, categorize it, and determine the frequencies. Despite this overall lack of guidance, methods for analyzing qualitative reactions do exist and are commonly used by qualitative researchers.

Because reactionnaires are a data collection instrument used to collect both quantitative and qualitative data, one can look to the areas of quantitative and qualitative research for guidance in analyzing the data. When analyzing quantitative data, one can reference many books on the subject, all covering the same statistical methods, and although the content is presented in different manners, a researcher who has selected an appropriate statistical test relative to the research being conducted can conduct that test accordingly and receive unquestionable results, for the statistical test is established (Tesch, 2013). However, many qualitative research practices are not as standardized and concrete as quantitative practices, and the term *qualitative research* itself is defined and understood differently by researchers in different fields (Coffey & Atkinson, 1996; Tesch, 2013; Silverman, 2015). Tesch (2013) identified 46 types of qualitative research

defined by researchers in social sciences, some of which overlapped and others that were not conceptually aligned. She noted that some researchers used a specific term to indicate the research perspective while others used the same term to represent the research method (e.g., grounded theory). According to Tesch, regardless of how qualitative research is approached, defined, and understood, "the only agreement we would find among qualitative researchers is that analysis is the process of making sense of narrative data" (p. 3).

Just as approaches to qualitative research are abundant, so too are approaches for analyzing qualitative data. For example, Tesch (2013) identified 26 different strategies for analyzing qualitative data. Furthermore, just as the term qualitative research has many meanings, so does the term *analysis* (Coffey & Atkinson, 1996). Nevertheless, however researchers approach, define, and understand QDA, they agree that there is no correct way to perform this analysis and that one must choose the analysis method that best fits the use case (Coffey & Atkinson, 1996; Miles, Huberman, & Saldaña, 2019; Patton, 2015).

Conceptual and Theoretical Frameworks for Describing the Phenomenon

This study approached the research problem from a human performance technology (HPT) perspective. According to the International Society for Performance Improvement, HPT is:

> A systematic approach to improving productivity and competence, uses [sic] a set of methods and procedures—and a strategy for solving problems—for realizing opportunities related to the performance of people. More specific [sic], it is a process of selection, analysis, design, development, implementation, and

> evaluation of programs to most cost-effectively influence human behavior and accomplishment. It is a systematic combination of three fundamental processes: performance analysis, cause analysis, and intervention selection, and can be applied to individuals, small groups, and large organizations. (J.A. Pershing, 2006, p. 9)

While HPT involves a systematic, comprehensive approach, this study centered only on performance analysis and cause analysis. HPT practitioners (also known as human performance improvement [HPI] practitioners) use performance analysis to identify the current state of performance, the desired state of performance, and any gaps between these states. HPT practitioners then use cause analysis to uncover the reasons for the gaps. This study focused on describing the current state of performance and the underlying factors that impact that performance as it relates to instructional designers' analysis of qualitative reactions.

A conceptual framework for describing performance analysis. Given the broad range of QDA methods, frameworks can help provide guidelines (Patton, 2015), and in the case of the current study, helpful in understanding and describing a phenomenon. Two of the most popular QDA frameworks cited in the literature were combined in this study to form the conceptual framework for describing instructional designers' current state of performance. The first QDA framework is from Yin (2015), who presented a five-phased approach: *compiling*, *disassembling*, *reassembling*, *interpreting*, and *concluding.* The second framework is from Miles et al. (2019), who reduced analysis to three activities: *data condensation*, *data display*, and *drawing and verifying conclusions*. While Yin's approach includes phases that map to the activities of

data condensation and drawing and verifying conclusions, his approach does not expressly convey data display. Therefore, Figure 2 displays the combined conceptual frameworks that provide a means by which one can categorize, classify, and discuss instructional designers' QDA procedures and the output they yield. This framework is underscored by its flexibility: instructional designers can move through each phase in a non-linear, iterative fashion, as needed.

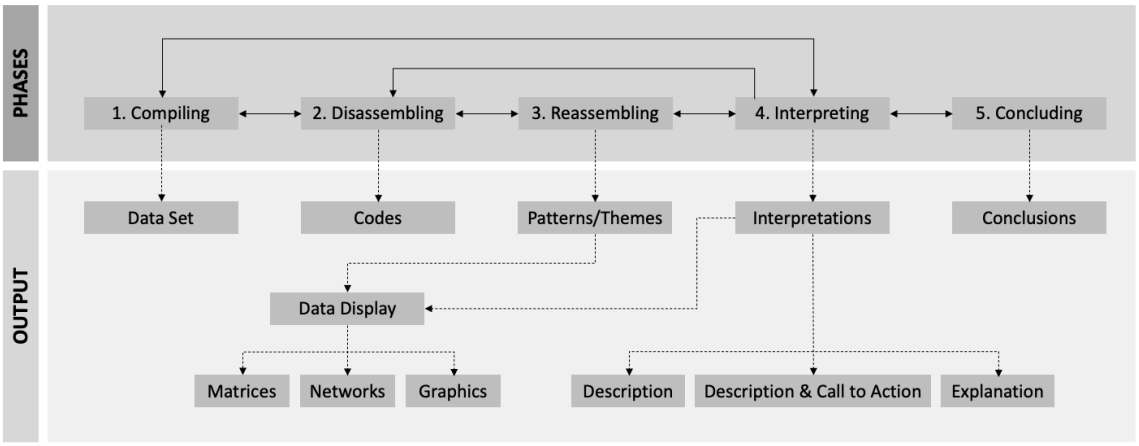


Figure 2. A conceptual framework for describing how instructional designers analyze qualitative reactions. The top portion of the figure incorporates Yin's (2015) five-phased approach to qualitative data analysis (QDA), while the bottom portion outlines the outputs of those five phases. The outputs incorporate the types of data displays as outlined by Miles, Huberman, and Saldaña (2018), as well as the types of interpretations outlined by Yin (2015).

The compiling phase. In the compiling phase, the analyst gathers and readies the data for disassembling. In this phase and subsequent phases, it is helpful to adopt the

data nomenclature that Braun and Clarke (2006) used in their outline of thematic analysis:

> Data corpus refers to all data collected for a particular research project… data set refers to all the data from the corpus that are being used for a particular analysis…. data item is used to refer to each individual piece of data collected, which together make up the data set or corpus.… data extract refers to an individual coded chunk of data, which has been identified within, and extracted from, a data item. (p. 79)

Thus, the output or product of the compiling phase is the data set that will be further analyzed. Concerning reactionnaires, the data set consists of all qualitative reactions across all reactionnaires, with each reactionnaire itself considered a data item and all reactionnaires forming the data corpus (see Figure 3).

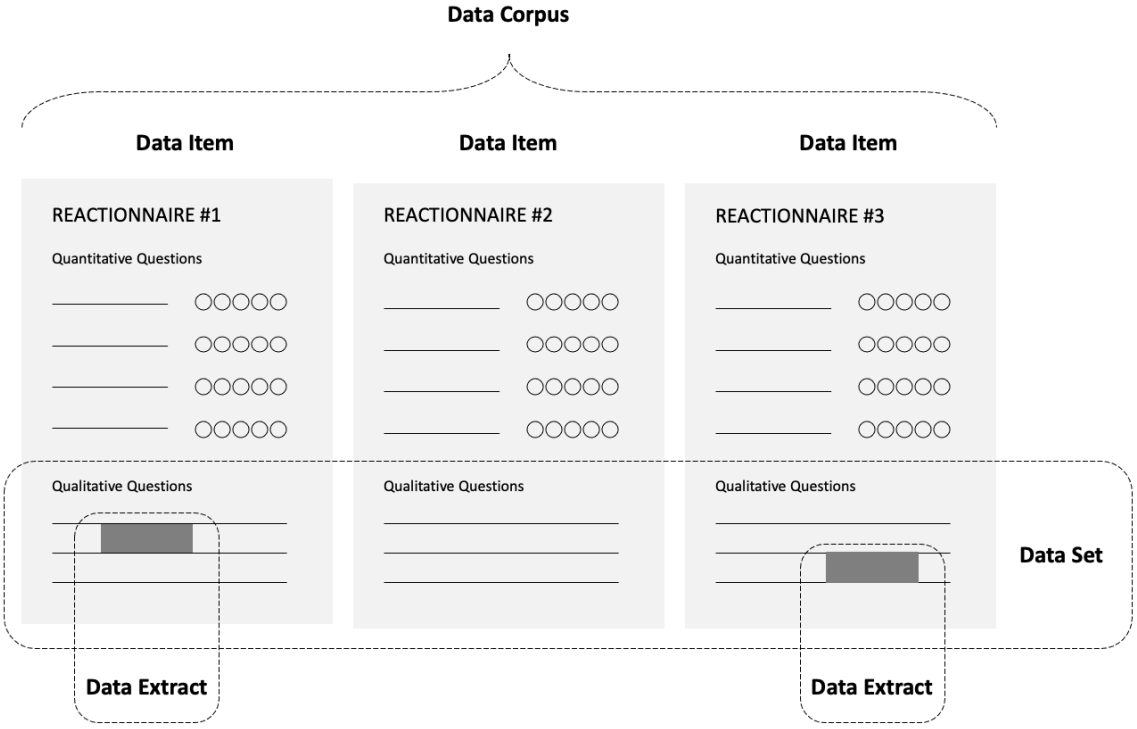


Figure 3. Braun and Clarke's (2006) data nomenclature as applied to reactionnaires.

The disassembling and reassembling phases. Fundamentally, the disassembling and reassembling phases are about deconstructing the data to find patterns and themes. While there are no fixed rules for disassembling and reassembling data, coding is considered a fundamental activity in QDA (Auerbach & Silverstein, 2003; Creswell & Creswell, 2018; Grbich, 2007; Tesch, 2013; Miles et al., 2019; Saldaña, 2015). A code is "most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data" (Saldaña, 2015, p. 4). These words and phrases can range from descriptive to abstract and come from the data itself or the analyst's vocabulary (Silverman, 2015). Researchers can develop predetermined codes that are applied to the data, develop codes as they emerge from the data, or use a combination of the two approaches (Creswell & Creswell, 2018).

Although there are many ways to code data, researchers are careful not to identify a single method as the established standard (Patton, 2015; Saldaña, 2015; Tesch, 2013). In examining characteristics of analysis across a broad range of research methods and practices, Tesch (2013) concluded, among other things, that "the process is systematic and comprehensive, but not rigid" (p. 95) and that "the procedures are neither 'scientific' nor 'mechanistic'" (p. 96). These statements mirror the literature on coding methods, which range from simple guidelines to more complete taxonomies. For example, Creswell and Creswell (2018) provided streamlined instruction on coding, specifying that

researchers should code based on topics they would expect to find, surprising findings that could not be anticipated in advance, and unusual or conceptually interesting ideas. In contrast, Saldaña (2015) presented a more comprehensive overview of coding methods, dividing 33 different methods into two cycles: *first cycle coding* and *second cycle coding*. First cycle coding is the process of assigning initial codes to the data, while second cycle coding is the process of developing "a smaller and more select list of broader categories, themes, concepts, and/or assertions" (p. 233). In other words, first cycle coding is the process of reducing the raw data to a set of codes, and second cycle coding is the process of reducing those codes to a smaller set. Thus, in the conceptual model displayed in Figure 2, the disassembling phase can be considered synonymous with first cycle coding and the reassembling phase with second cycle coding.

Computer-assisted qualitative data analysis software (CAQDAS). Researchers historically coded data "by hand," using tools such as pens, pencils, highlighters, and sticky notes, among other tools (Hart & Achterman, 2017). While hand-coding text is still a primary method used by researchers, it can be an arduous task, and researchers often use CAQDAS to assist with the process (Creswell & Creswell, 2018). However, CAQDAS can only assist with the analysis process and does not analyze the data, for that is a task that must be completed by the researcher (Hart & Achterman, 2017; Maxwell, 2018; Miles et al., 2019; Patton, 2015; Yin, 2015). While many CAQDAS programs exist, some of the most popular programs include ATLAS.ti, MAXQDA, and NVivo (Castleberry & Nolen, 2018; Creswell & Creswell, 2018; Hart & Achterman, 2017; Verdinelli & Scagnoli, 2013). These programs allow researchers to perform important coding tasks, such as importing predetermined codes that will be used in the coding

process; creating new codes throughout the process; revising codes throughout the process; adding notes, memos, and files to codes; and running searches and reports on the data (Hart & Achterman, 2017).

The interpreting and concluding phases. According to Yin (2015), "interpreting may be considered the craft of giving your own meaning to your findings" (p. 220). Because interpretation is fundamentally subjective, Yin offered three broad types of interpretation that we can use for classification purposes: *description*, *description plus a call for action*, and *explanation*. At its core, description refers to the act of describing the themes and patterns uncovered in the data, while description plus a call for action incorporates the prompting of a follow-up activity; explanation focuses on "explaining how or why things happened, or alternatively how or why people said what they did" (p. 231). Although the interpreting phase may appear to be a terminal phase in QDA, one more phase—the concluding phase—is where the analyst forms conclusions that tie back to the purpose of the analysis itself. In other words, the concluding phase is where the analyst answers the question that prompted the need for the QDA.

Data display as a fundamental activity in QDA. In QDA, an analyst may develop data displays during the reassembling and interpreting phases to visualize the data. Miles et al. (2019) defined data display as "an organized, condensed assembly of information that allows analytic reflection and action" (p. 8) and presented three groupings of data displays: *matrices* that organize data by rows and columns; *networks* that use lines and arrows to connect data points; and *graphics*, which serves as a catch-all for a wide assortment of other designs and images. While data displays can be helpful in effectively depicting information and can generally be used for final reporting of the data

(Miles et al., 2019), they can also present challenges. One major challenge is that basic displays do not highlight connections among the data, though this can be lessened by writing accompanying narratives that include inferences about any relationships (Slone, 2009; Williamson & Long, 2005). A second major challenge is that the development of displays can be time-consuming, and one may not have the technology, technical aptitude, or both to create more sophisticated displays (Henderson & Segal, 2013; Miles et al., 2019; Slone, 2009; Williamson & Long, 2005). Additionally, related to this challenge, some technologies may be easy to use but generate displays that are less informative than other displays (Henderson & Segal, 2013). Thirdly, the development of data displays is dependent on the analyst's creative abilities, for displays can evolve throughout the analysis process (Williamson & Long, 2005).

A theoretical framework for describing cause analysis. While this study used the conceptual model displayed in Figure 2 to describe instructional designers' current state of performance as it relates to the analysis of qualitative reactions, the study also used Chevalier's (2003) update of Gilbert's (2007) Behavior Engineering Model (BEM) to describe the underlying factors impacting instructional designers' current performance. BEM "provides a framework with which we may analyze the different environmental and individual elements that come together and influence performance in any work setting" (Austin & Garnier, 1998, p. 13). Figure 4 displays Gilbert's (2007) original BEM, and Figure 5 displays Chevalier's (2003) updated BEM, which includes six core performance factors categorized as either environmental or individual. In the updated model, Chevalier revised some of Gilbert's original terminology and recommended that HPT practitioners begin cause analysis with the three environmental factors: *information*,

resources, and *incentives*. These environmental factors "pose the greatest barriers to exemplary performance… are easier to improve and have a greater impact on individual and group performance" (Chevalier, 2003, p. 4). Chevalier advanced this standpoint by arguing that even if one were to change the individual factors of *knowledge/skills*, *capacity*, and *motives*, which are more difficult to change, "performance will most likely not improve if there are environmental factors that remain unsolved" (Chevalier, 2003, p. 6).

	Information	**Instrumentation**	**Motivation**
Environmental Supports	**Data** 1. Relevant and frequent feedback about the adequacy of performance 2. Descriptions of what is expected of performance 3. Clear and relevant guides to adequate performance	**Instruments** 1. Tools and materials of work designed scientifically to match human factors	**Incentives** 1. Adequate financial incentives made contingent upon performance 2. Nonmonetary incentives made available 3. Career-development opportunities
Person's Repertory of Behavior	**Knowledge** 1. Scientifically designed training that matches the requirements of exemplary performance 2. Placement	**Capacity** 1. Flexible scheduling of performance to match peak capacity 2. Prosthesis 3. Physical shaping 4. Adaptation 5. Selection	**Motives** 1. Assessment of people's motives to work 2. Recruitment of people to match the realities of the situation

Figure 4. Behavior Engineering Model. From *Human Competence: Engineering Worthy Performance* (p. 88), by T.F. Gilbert, 2007, San Francisco, CA: Pfeiffer. Copyright 2007 by the International Society for Performance Improvement.

Environment	**Information** 1. Roles and performance expectations are clearly defined; employees are given relevant and frequent feedback about the adequacy of performance. 2. Clear and relevant guides are used to describe the work process. 3. The performance management system guides employee performance and development.	**Resources** 1. Materials, tools and time needed to do the job are present. 2. Processes and procedures are clearly defined and enhance individual performance if followed. 3. Overall physical and psychological work environment contributes to improved performance; work conditions are safe, clean, organized, and conducive to performance.	**Incentives** 1. Financial and non- financial incentives are present; measurement and reward systems reinforce positive performance. 2. Jobs are enriched to allow for fulfillment of employee needs. 3. Overall work environment is positive, where employees believe they have an opportunity to succeed; career development opportunities are present.
Individual	**Knowledge/Skills** 1. Employees have the necessary knowledge, experience and skills to do the desired behaviors 2. Employees with the necessary knowledge, experience and skills are properly placed to use and share what they know. 3. Employees are cross- trained to understand each other's roles.	**Capacity** 1. Employees have the capacity to learn and do what is needed to perform successfully. 2. Employees are recruited and selected to match the realities of the work situation. 3. Employees are free of emotional limitations that would interfere with their performance.	**Motives** 1. Motives of employees are aligned with the work and the work environment. 2. Employees desire to perform the required jobs. 3. Employees are recruited and selected to match the realities of the work situation.

Figure 5. Updated Behavior Engineering Model. From "Updating the Behavior Engineering Model," by R. Chevalier, 2003, *Performance Improvement*, *42*, p. 3. Copyright 2002 by Roger D. Chevalier.

Combining the conceptual and theoretical frameworks. Figure 6 displays the combined conceptual and theoretical frameworks that form the underlying structure used to describe how instructional designers analyze qualitative reactions. The top portion of the figure, the conceptual framework for performance analysis, displays the phases and output of QDA. The bottom portion of the figure, the theoretical framework for cause analysis, displays the BEM. The dotted line that connects the bottom portion of the figure to the top illustrates how instructional designers' analysis of qualitative reactions (i.e., performance) is impacted by underlying environmental and individual factors (i.e., causes). Thus, the performance analysis framework was used to describe the current state

of performance of instructional designers as it relates to their analysis of qualitative reactions, while the cause analysis framework was used to describe the underlying reasons for the current state of performance.

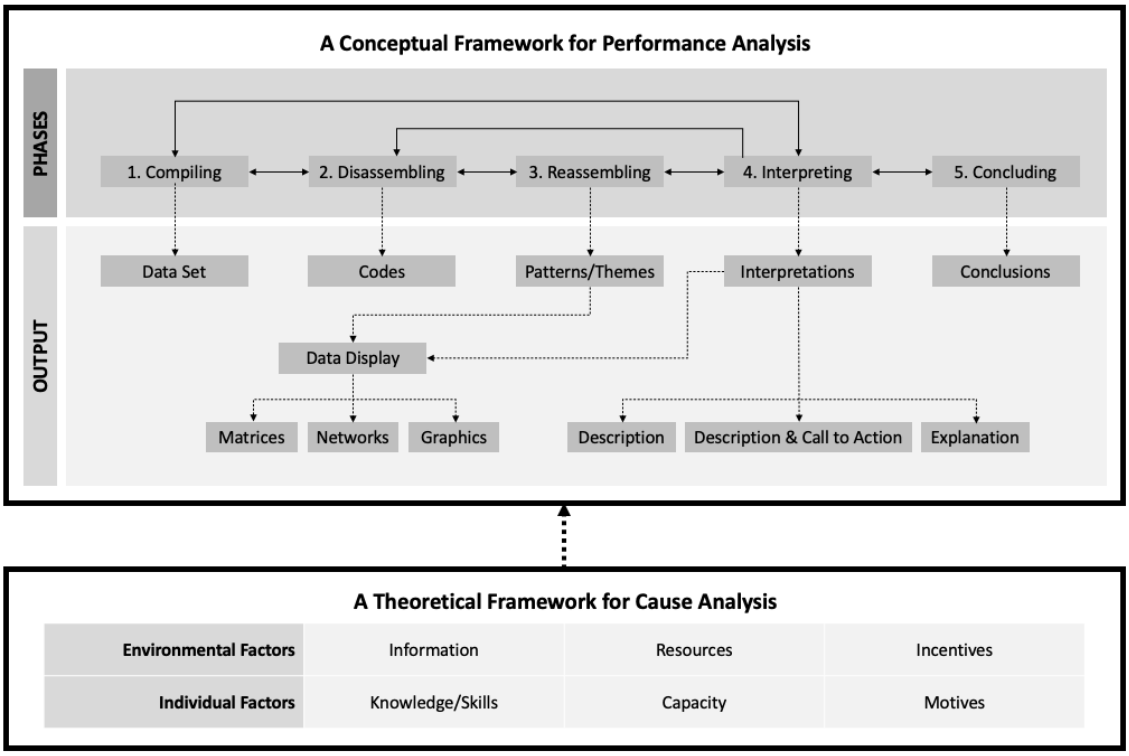


Figure 6. Underlying structure for describing the phenomenon. The conceptual framework for performance analysis incorporates qualitative data analysis (QDA) concepts from Miles et al. (2018) and Yin (2015). The theoretical framework for cause analysis is drawn from Chevalier's (2002) updated Behavior Engineering Model.

Summary

It has been more than 60 years since the Kirkpatrick model was first introduced, and although researchers question the model's soundness, the Kirkpatrick model remains

the most popular training evaluation model. However, researchers also acknowledge that instructional designers' use of the model is frequently limited to Level 1. Still, evaluation at this level produces data that are used for an array of decision-making purposes surrounding the design, development, delivery, and improvement of training programs.

The primary means by which instructional designers conduct Level 1 evaluation is via reactionnaires, which measure a range of dimensions using a broad mix of closed- and open-ended questions. Given Donald L. Kirkpatrick's initial guidance that reactionnaires should be designed so they can be easily tabulated, it is not surprising that closed-ended questions are the most popular question type found in reactionnaires. Researchers have confirmed that these questions regularly use Likert scale question responses that are converted to numbers and then averaged, thus adhering to Kirkpatrick's tabulation guideline. Researchers also agree that open-ended questions are valuable and should be used in reactionnaires, yet despite this assertion, relatively little guidance for instructional designers is available in the literature on analyzing qualitative reactions; however, a body of knowledge on this subject does exist for researchers.

Bates (2004) contended that it is an ethical responsibility for scholars to research training evaluation. Giangreco et al. (2010) agreed and pointed to the scant body of literature on Level 1 evaluation itself, arguing that academics must dedicate more research to the actual types of evaluations being used. Thus, given the facts that reactionnaires are the most popular means by which instructional designers gather data to make training program decisions, that instructional designers include in their reactionnaires open-ended questions that yield crucial feedback, that researchers encourage this practice, and that there is a lack of guidance for instructional designers on

analyzing qualitative data, there was a need to investigate how instructional designers analyze qualitative reactions. Chapter three details the methods used to conduct this investigation.

Chapter 3

Methods

This study sought to uncover how instructional designers analyze qualitative trainee reactions and the underlying reasons for that phenomenon. This chapter details the methods used to carry out the research. The chapter begins with an overview of the research design, followed by a description of the population and the criteria used in selecting the study participants. An interview protocol is then described, as are the procedures used to perform and record the interviews. Next, the methods used to analyze and synthesize the collected data are described and supported by a description of the processes used to ensure reliability and trustworthiness and an explanation of the researcher's role. Finally, the study's limitations are presented.

Research Design

Pragmatism as a research paradigm emphasizes the research problem and is "oriented toward 'what works' and real-world practice" (Creswell & Plano Clark, 2018, p. 37). Guided by pragmatism, the researcher viewed the research problem from the lens of a human performance improvement practitioner, seeking to uncover the participants' current state of performance regarding the analysis of qualitative trainee reactions and the underlying causes for that state of performance. Given that performance and cause analyses are exploratory activities, an overall qualitative research approach was selected, which Creswell and Creswell (2018) deem appropriate for exploratory research that is often prompted by a lack of existing research on the topic. More specifically, a phenomenological methodology was employed, which is descriptive in nature, "emphasizing hermeneutic or interpretive analyses of lived experiences" (Yin, 2015, p.

20). The phenomenological methodology used semi-structured interviews to uncover participants' lived experiences with the phenomenon. Figure 7 displays the overall research design.

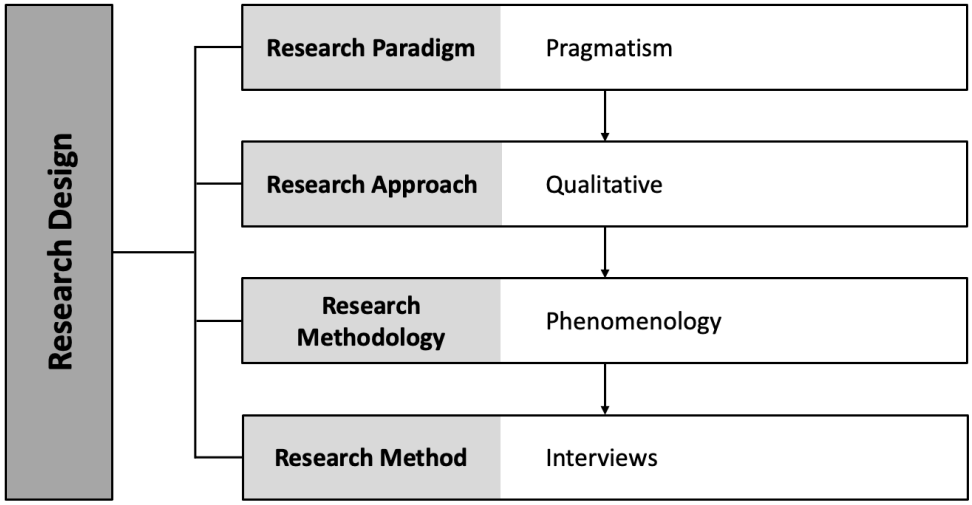


Figure 7. The current study's research design.

Setting

This research was carried out via semi-structured interviews using Zoom, a video communications platform. The setting was selected for two primary reasons. First, the participants were located across the United States, making virtual interviews a more practical method than face-to-face interviews that would have required travel. Additionally, the COVID-19 pandemic limited the researcher's and participants' abilities to travel and meet face-to-face. Second, it was assumed that most participants were familiar with video communications platforms because they had used them before or during the COVID-19 pandemic for work- or training-related purposes.

Sampling Procedures

The study's population consisted of instructional designers who collected qualitative reactions. As it relates to phenomenological research, Creswell and Poth (2016) state that "it is essential that all participants have experience of the phenomenon being studied" (p.156). Therefore, a purposive, criterion-based sampling strategy was used to select a sample of the population who could help answer the research questions. To be eligible to participate in the study, members of the population were required to have:

- Provided consent to participate in the study.
- Held a job that involved the design, development, implementation, evaluation, or management of training.
- Offered at least two training programs to their learners in the past 6 months.
    - Evaluated the training program(s) using a reactionnaire that included at least one qualitative question.

A recruitment survey was created using these criteria (see Appendix A) to identify suitable participants for the study. The survey was posted on the researcher's LinkedIn newsfeed (see Appendix B), where it was reshared by the researcher's network of peers. The recruitment survey resulted in 148 respondents, of which 13 were selected to participate in the study. The selection process involved ensuring the participants met the required criteria and that there was maximum variation so that the group of participants reflected the diversity of the population and could help reach data saturation.

Instruments

Creswell and Poth (2016) indicate that the researcher is a primary research instrument in qualitative research and typically develops any associated instruments. In the current study, an interview protocol (see Appendix C) was created using the researcher's existing knowledge of and experience with training program evaluation practices, a review of literature that was pertinent to this study, and the conceptual and theoretical frameworks in Figure 6. Given that participants would be focusing on past experiences, the questions were written and ordered to help stimulate a recall of those previous experiences. The interview protocol included four demographic questions (DQ) and seven interview questions (IQ), as displayed in Table 1.

Table 1

Semi-structured Interview Questions

Question	Rationale for Question
DQ1: Would you mind sharing a bit about the organization you work for and your role at that organization?	To gather demographics and establish rapport.
DQ2: Would you mind sharing a bit more about your background in the learning and development field and your background evaluating training?	To gather demographics and establish rapport.
DQ3: Tell me about the training programs you offered over the past 6 months.	To gather demographics and establish rapport.
DQ4: Tell me about the open-ended questions you included in your reactionnaires.	To gather demographics and establish rapport.
IQ1: After each training program ended and you collected the reactionnaires, walk me through the steps you took to analyze the open-ended responses.	To answer RQ1.
IQ2: Tell me a bit about why you analyze open-ended responses the way you do.	To answer RQ2.
IQ3: What challenges do you face when analyzing open-ended responses?	To answer RQ2.
IQ4: What support do you receive when analyzing open-ended responses?	To answer RQ2.
IQ5: What do you think your personal strengths are when analyzing open-ended responses?	To answer RQ2.
IQ6: What motivates you to perform the analysis?	To answer RQ2.
IQ7: What do you think would help you perform your analysis better?	To answer RQ2.

Data Collection Procedures

An Institutional Review Board (IRB) request to Baker University was submitted on March 12, 2021 (see Appendix D) and approved on March 23, 2021 (see Appendix E). On April 22, 2021, the researcher notified the IRB that the protocol was revised to include compensation in the form of a $25 Amazon gift card for individuals who completed an interview; the IRB indicated that this was acceptable. The research participation request in Appendix B was posted to the researcher's LinkedIn newsfeed on April 25, 2021, to target members of the population and identify suitable interview candidates. The post included a link to the criterion-based survey in Appendix A. If a participant clicked the survey link, the survey launched in a browser window via an HTTPS (Hypertext Transfer Protocol Secure) connection. After the participant reviewed a welcome message and an overview of the study's background information, an informed consent notice was displayed. To continue with the survey, the participant had to provide consent; if they did not consent, they were exited from the survey. Once the participant completed the survey, a thank you message was displayed. Their responses were recorded and stored in SurveyMonkey, which was only accessible to the researcher via his login credentials. All responses remained confidential throughout the data collection process.

Thirteen participants were selected to interview and were notified via email (see Appendix F). The email notification included a copy of the interview questions and an interview consent form, which participants were asked to review, sign, and return before the interview. All interviews occurred via Zoom between June 1, 2021, and July 20, 2021, and lasted no more than 60 minutes. When the interview occurred, several steps

were taken before asking the interview questions. First, the study's purpose was reshared. Second, receipt of the signed interview consent form was confirmed. Third, the participant was informed that they could choose to stop the interview at any point and could also choose not to answer any questions. Fourth, the participant was informed that the interview would be recorded. Fifth, the participant was reassured that their identity and the interview would remain confidential. Sixth, the participant was allowed to ask any clarifying questions about the interview process. Finally, the participant was asked to provide verbal consent to record the interview. Once the permission was obtained, the recording was started in Zoom. The recording was stored in Zoom's password-protected cloud environment and later deleted when the study concluded. Following the interview, the participant received a $25 Amazon gift card that included a thank you message.

Data Analysis and Synthesis

The following research questions guided this study:

RQ1: How do instructional designers analyze qualitative reactions?

RQ2: What rationale do instructional designers provide for the methods they use to analyze qualitative reactions?

Because a conceptual framework for QDA had been created to help describe the phenomenon (see Figure 2), the same framework was used to analyze the interview data. To compile the data, Zoom was used to produce transcriptions of each recorded interview. Once Zoom generated a transcript, it was saved as a text file and imported into MAXQDA, where it was reviewed against the interview recording and corrected where necessary. Additionally, the participant's name was coded to maintain anonymity (e.g., Participant A, Participant B, etc.); the participant coding matrix was stored in a password-

protected electronic format that was only accessible to the researcher and was permanently deleted once the study concluded.

To disassemble the data, each transcript was read line-by-line in MAXQDA. As the transcripts were read, initial codes were assigned to words and phrases. The specific codes that were assigned represented the overall essence of the text. While pre-determined codes were sourced from Figure 2, codes were also developed as they emerged. Once all initial codes were assigned, the data was reassembled by reducing the initial codes to a set of patterns and themes. Matrices and network displays were created to assist with the reassembly process. Both the disassembling and the reassembling phases were iterative, requiring continual re-reading and re-coding of the data.

The interpretation phase began after the data was reassembled. Since the study was exploratory and descriptive, a descriptive interpretation (Yin, 2015) was used to describe the themes and patterns uncovered in the data. All five of Yin's criteria for comprehensiveness were considered when forming the interpretation: *completeness*, *fairness*, *empirical accuracy*, *value-added*, and *credibility*. Finally, the results of the interpretation were used to draw conclusions, answer the research questions, and suggest additional research. The results of the interpretation and conclusion phases are presented across Chapters 4 and 5.

Reliability and Trustworthiness

The merit of a qualitative study depends on the measures used to establish and maintain reliability and trustworthiness (Creswell & Plano Clark, 2018). In qualitative research, trustworthiness is tied to four key criteria defined by Lincoln and Guba (1985): *credibility*, *transferability*, *dependability*, and *confirmability*. To guarantee the reliability

and trustworthiness of the study, five techniques driven by Lincoln and Guba's criteria were incorporated. First, an impartial colleague of the researcher was selected to perform peer debriefing (Creswell & Creswell, 2018; Lincoln & Guba, 1985). The colleague was provided with the study's background and a copy of the interview protocol. Based on feedback provided by the colleague, several interview questions were reworded for clarity. Second, after the interview protocol was revised, two pilot interviews were conducted with individuals from the researcher's professional network. As a result of the pilot interviews, some interview questions were further refined, and two question probes were added to the protocol. Third, a consistent approach was maintained in each interview using Creswell and Poth's (2016) interview guidance: "Stay within the study boundaries you have reviewed, use the protocol to guide your questions, complete the interview within the time specified, be respectful and courteous, and offer few questions and advice" (p. 165). Additionally, each interview transcription was carefully reviewed against the interview recording, and any errors were corrected before the collective data was analyzed. The transcription review process generated clean verbatim transcriptions—precise transcriptions of what the participants said with light editing for readability (i.e., filler words, repeated words, false starts, and other distractions were removed). Fourth, thick descriptions (Creswell & Poth, 2016) were used to describe the participants and their emerging themes. Lastly, as evidenced in this chapter, all processes used to carry out the research were documented.

Researcher's Role

In qualitative research, a researcher's "personal background, culture, and experiences hold potential for shaping their interpretations, such as the themes they

advance and the meaning they ascribe to the data" (Creswell & Creswell, 2018, p. 278). This study's researcher holds an M.Ed. in instructional design, is a Certified Professional in Talent Development (CPTD), and has worked in the instructional design and performance technology field for roughly 13 years, primarily in leadership positions at for-profit companies. His interest in studying the phenomenon was related to his own experience in evaluating training programs. To minimize bias throughout the study, the researcher:

- Incorporated the methods described in the previous section to guarantee the overall reliability and trustworthiness of the study.
- Bracketed (i.e., set aside) his personal experiences with the phenomenon to focus on those of the participants. To accomplish bracketing, the researcher reflected on his beliefs about, experiences with, and knowledge of the phenomenon, then identified how those factors might influence the study. The exercise prompted the researcher to ensure he consistently explained the study as clearly as possible to participants, omitting any personal opinions, thoughts, or ideas. He was also prompted to avoid influencing participants' understanding of the phenomenon by allowing them to describe their experiences with the phenomenon using their own words. While he developed the frameworks in Figure 6 before the interviews occurred, he only used them after all interviews were completed to analyze and describe the data via standardized nomenclature.
- Asked participants a consistent set of factual and probing questions that were relevant to the research questions.

- Discussed the study on an ongoing basis with his major advisor and research analyst.

Limitations

According to Lunenburg and Irby (2008), a study's limitations are "factors that may have an effect on the interpretation of the findings or on the generalizability of the results" (p. 133). This study had four limitations. First, although the researcher aimed for maximum variation in sampling, and given the small sample size, it is possible that the participants were not fully representative of the population. Second, despite the researcher's best effort to select participants who could help answer the research questions, it is possible that the group of participants who were not selected for interviews included individuals who could have provided additional information outside of that provided by the participants. As a third limitation, the COVID-19 pandemic restricted participants from delivering live, face-to-face training, which, in turn, restricted them from collecting paper-based reactionnaires. Before the pandemic, the use of paper-based reactionnaires was a well-established practice for instructional designers. Thus, the participants' experiences with the phenomenon were limited to the use of electronic reactionnaires. Fourth, it is assumed the participants understood all interview questions and provided unbiased responses; however, the information provided by participants could not be independently verified.

Summary

Two research questions guided this phenomenological research study. To answer the research questions, semi-structured interviews were conducted with 13 participants. The interviews were conducted using Zoom and guided by an interview protocol that

included 11 questions. Transcriptions of each interview were uploaded to MAXQDA, and the data was analyzed using the conceptual framework for QDA that was created to help describe the phenomenon. Chapter 4 presents the results of the analysis.

Chapter 4

Results

There were two purposes for this study. The first purpose was to examine the methods used by instructional designers to analyze qualitative reactions. The second purpose was to examine the underlying reasons for the use of those methods. To carry out the study, semi-structured interviews were conducted with 13 participants. This chapter presents the results of the analysis of the interviews. The chapter begins with a description of the participant demographics. Next, the training programs that the participants evaluated via reactionnaires are described. Finally, the themes that emerged for Research Questions 1 and 2 are presented.

Participant Demographics

The participants' ages ranged from 24 to 51, with the median age being 40. Eleven participants identified as female, and two identified as male. All participants resided in the United States. Eleven participants held at least one master's degree, one participant held a bachelor's degree, and one held multiple technical certifications. The participants had between 1 and 30 years of experience evaluating training, with the median number of years of experience being 12. Some participants indicated that the extent of their experience with training evaluation was limited to what they were doing with their current employer, while other participants indicated they had evaluated training in previous jobs and throughout their careers.

Eleven participants worked at for-profit organizations across a range of industries. These organizations employed roughly 120 to 100,000 employees. One participant worked for a government entity that employed approximately 45,000 employees, while

another participant was an independent learning and development consultant. Although the participants were employed in various roles that all shared the responsibility of training evaluation, most participants reported that their roles also involved the design, development, and delivery of training. Moreover, all but one participant had delivered the training programs they had evaluated via reactionnaires. Table 2 summarizes the participant demographics.

Table 2

Participant Demographics

Participant ID	Age	Gender	State of Residence	Highest Level of Education/Degree Earned	Years of Experience Evaluating Training (Approximate)	Description of Current Employer (Employment Numbers are Approximate)
A	38	Female	NY	Master's in Higher Education	6	Technology distribution company; employs 40,000 employees.
B	31	Female	MA	Master's in International Education	7	Online brokerage company selling small business insurance; employs 1,000 employees.
C	34	Female	TX	Master's in Human Resource Management	6	Mortgage company; employs 1,000 employees.
D	37	Female	NY	Master's in Elementary Education (Grades 1-6)	20	Telecommunications and mass media company; employs 100,000 employees.
E	51	Male	MI	Master's in Business Administration and Master's in Human Resources Development	25	Direct selling and network marketing company; employs 16,000 employees.
F	51	Female	MA	Master's in Education (Learning Psychology and Instructional Technology)	30	Independent learning and development consultant.

G	24	Female	MA	Bachelor of Science (Economics and Applied Psychology)	1	e-Commerce company selling furniture and home goods; employs 15,000 employees.
H	43	Male	MA	Master's in Organizational Learning and Development	15	Biopharmaceutical company; employs 1,500 employees.
I	36	Female	SC	Multiple Technical Certifications	11	Government entity responsible for public security; employs 45,000 employees.
J	45	Female	UT	Master's in Adult Education and Training	19	Telecommunications company; employs 120 employees.
K	39	Female	CT	Master’s in Business Administration	4	Personal and business insurance company; employs 30,000 employees.
L	38	Female	NY	Master’s in Business Administration	2.5	Regional bank; employs 17,000 employees.
M	49	Female	NJ	Master’s in Business Administration	10	Engineering and industrial software company; employs 6,000 employees.

Description of Training Programs Evaluated via Reactionnaires

Ten participants reported that the programs they evaluated via reactionnaires were delivered to internal audiences (i.e., employees), some of which were globally dispersed. Three participants reported that their programs were delivered to both internal and external audiences (i.e., non-employees, such as product users or vendors). All participants reported that because of the COVID-19 pandemic, their programs were delivered via live, virtual methods. Some programs also used a blended approach that combined live, virtual training with eLearning and other training methods and materials. While the participants reported that they evaluated various training programs, most programs were new hire orientation programs, leadership training programs, or both. Additional training programs that were evaluated focused on diversity and inclusion, products, sales, soft skills, and systems. The length of the training programs spanned from less than 1 hour to several months and included anywhere between three and 60 individuals per instance.

Emerging Themes from Research Question 1

Research Question 1 sought to uncover how instructional designers analyze qualitative reactions. The responses to the interview questions for RQ1 revealed that the participants used a variety of processes and methods to analyze their data sets. For example, Participant J reported that her process involved "taking a brief look, seeing if there's anything that needs to be improved," while Participant D described how her method was "very anecdotal" and involved a team that "message[ed] each other, 'Hey this person just said this. What do you think about that?' Or, 'Hey, this is good news.' A lot of it's that way." As another example, Participant F described how her process was

"casual in the sense that I'm reading through them and then I just note what I can take action on." Participant I used the same approach as Participant F but also entered her notes into a cumulative spreadsheet so she could spot trends over time. Participant L described how she searched for constructive feedback as a starting point:

> I immediately look for anything where it was like strongly disagree… and sometimes, when I see a strongly disagree and I read the comments, I'm like, "Oh, I think that this person just clicked the wrong number cause the comments support a different answer." So yeah, I look for the constructive stuff, so like, what can I do better. And then if I see it, then I read it. And if I don't, then I go through the full list, and I start top to bottom.

As a final example, Participant B described her color-coding approach to identifying trends:

> I'm a color-coded freak by nature… I'll take a look through—just first look—and then I'll go in and based off that, I will color-code those keywords that I'm seeing, and that's kind of what helps me gather my top five or top three themes from the open-ended questions.

Given the participants' various processes and methods to analyze their data sets, three themes emerged from the interview data: surveying platforms; keywords, patterns, and themes; and team effort.

Surveying platforms. To compile the data set (see Figure 3 for data nomenclature), all but one participant used at least one type of surveying platform to collect the qualitative reactions, with the most cited platforms being Survey Monkey, Microsoft Forms, and Qualtrics. Some participants exported the data set from the

surveying tool to Microsoft Excel for analysis, while others reviewed each data extract directly in the surveying tool. The decision to export the data set was dependent on the number of qualitative reactions received. For example, because of the high number of qualitative reactions she had to review, Participant A found it easier to work with the data in Excel, yet Participant H reviewed his qualitative reactions directly in Microsoft Forms because he did not have as many.

Keywords, patterns, and themes. Although the methods used to disassemble and reassemble the data set were mixed, most participants indicated they looked for keywords, patterns, or themes. To find the keywords, patterns, or themes, the participants used a variety of tactics. Some participants used more simplistic and unstructured approaches that included "kind of eyeballing it" (Participant K) and "just looking for words and themes that are the same" (Participant A), while others used more systematic processes, such as Participant G:

> As far as the qualitative feedback goes… looking for keywords there, so a lot of just splitting up the comments into columns [in Excel]… We'll go through and code all of them, just to say, like this is a networking comment, this is a manager comment… We'll read through them quickly. There's a lot of comments to go through, so we try to capture basically the gist of their comment into one specific category. And usually, the comments tend to fall into specific categories, so we'll try to cite those, and then we'll have another miscellaneous category, and then we'll look through miscellaneous comments later on.

Team effort. Most participants indicated they reviewed the data set as a team. There were two main reasons cited for the team effort. First, reviewing the data as a team

helped the individuals verify their analysis of the data. For example, Participant A described how reviewing the data as a team helped ensure the analysis remained objective:

> We would all go through it separately, pull out our themes, and then come together and see if our themes matched… because, of course, we have biases, right? So, we'd come together and see like, "Did I read it the same way you did?"… We all kind of come back with the same themes and make sure that we were hitting the mark on those and questioning ourselves and each other, too. Like I would say, "I don't know [coworker's name], I really saw that and like two comments, so while you think it's a theme, I really didn't." So, we kind of challenge each other on that, which was helpful because, again, we all have biases that we're reading with and the lens we're coming at, especially if it's the part that I facilitated or whatever. So, that was helpful to really kind of analyze it separately and then come together to talk about it and talk about those takeaways.

In another example, Participant B described how reviewing the data set as a team helped ensure the analysis was valid:

> What I do is I take out just overall themes of what I'm gathering, so like the top five themes based on the open-ended feedback. And then generally, after that, I'll have a team member take a look at it. I'll have them not only take a look at the five themes but also look at the raw data as well, just to kind of sense check if I'm missing anything or if they got something else out of it.

Second, reviewing the data set as a team provided a means to gather more context to better understand and interpret the qualitative reactions. Participant E noted that

including stakeholders such as trainee's managers can help "add some additional depth to those survey responses," while Participant M noted that her team discussed "any direct feedback that we got as instructors that we heard as we're talking to other people in the company that took the class."

Emerging Themes from Research Question 2

Research Question 2 sought to uncover the rationale behind the methods instructional designers use to analyze qualitative reactions. The responses to the interview questions for RQ2 revealed that the participants had a variety of underlying reasons for using the QDA processes they described. Given the variety of underlying reasons, three themes emerged from the interview data: primary motives, varied competencies, and support systems.

Primary motives. All participants reported motives for analyzing the data. The motives fell into three primary categories, and several participants shared multiple motives across these categories. The first category of motives was learner-centric motives, which were tied to the learners themselves. The participants who shared learner-centric motives emphasized that they wanted their training programs to be practical, helpful, and impactful for learners. For example, Participant F shared the following:

> I genuinely want to create something that is practical and helpful back on the job. And there's just no other way I'm gonna learn that. And so that's what motivates me. I'm passionate about not wasting anybody's time; you know what I mean? And so, I want to know if this actually created value for the participant and for the overall client if that's in the appropriate context. So, I'm just passionate about

> knowing about the quality of the experience. If it wasn't helpful, if it wasn't delivered in a quality way, then that's wasting people's time and resources, and I don't want that. I want people to truly be enabled to perform better or perform in a new way if that's what they want.

Participant D shared how she "want[s] things to be the best that we can make them at that time" and how she wanted to improve things for future learners. Some of the participants echoed Participant D's desire to improve things for learners. For example, Participant G said that "any improvements that we can make are always important," while Participant J shared the following: "I love to make things better for the person. I'm very much a—I guess people person or people pleaser, and I want to make sure that what I'm providing is the best to everybody."

The second category of motives was trainer-centric motives, which were tied to the trainers themselves, their performance, and their growth and development. Participant M shared how reviewing the data was "really about perfecting my craft," and Participant A emphasized how she used the data to identify areas of improvement for herself: "So, I'm always looking to see how I can improve personally, how I can improve my facilitation skills, how I can improve any content I'm delivering, how I can improve the way I'm asking questions." Furthermore, Participant I shared how the data was vital for her growth and development:

> When I'm instructing—well, with anything really, whether it's instruction, curriculum development—no one's perfect. I'm definitely not perfect, and I always want to continue to grow and develop myself, so I really want that feedback because I might be thinking I'm doing something really well, [but]

> maybe the way I'm delivering it is not well. Or maybe I wrote something… but it did not work out. Or, on the opposite side, maybe it did, and let's incorporate more of this type of activity into another course. So, I crave that feedback, and I think it's really imperative in my position in order to deliver the best product for instructors and for students.

The third category of motives was organization-centric motives, which were tied to the participants' departmental or organizational needs and goals. The participants who shared organization-centric motives focused on the value they brought to and the impact they had on their organizations. For example, Participant E said he was "motivated to do the work because I want to know and be able to show that we're making a business impact." Participant K shared that "like any other group, we have to provide value to the business to be here." Participant H elaborated on the importance of proving his team's value and how Level 1 evaluations played a role in that process:

> You know, I think that in a lot of cases we look like we are just a cost center, so in many cases we just spend money. We don't bring in money in terms of a training department, so we're always going out to vendors, spending money, and stuff like that. What we need to be able to show is that there's the value that our teams do bring to the organization. So, to me, the level ones just kind of start off that process.

Varied competencies. Most participants reported skills, knowledge, and abilities that came into play when analyzing their data sets. The collective competencies were mixed. For example, some participants shared how they could easily spot trends, such as Participant D:

> I am like the trend spotter. I can see things coming from a mile away. So, I can look at something that's happening and instantly recognize what's causing that issue, what we can do to fix that issue, how it's going to have a downstream impact or a long-term impact if you continue to do something that way. I am particularly good at that.

Participant I also noted how she could spot trends but added that she was detail-oriented: "I pay a lot of attention to detail, so it's easy for me to identify trends and really group and categorize the comments to easily input it into our tracking system." Participant C shared how she could easily find shortcomings: "I always say one of my strengths is I can walk into something and I can find a fault like that... I hate these jargony things... but getting into the weeds of really nitpicking things that may go unnoticed." Some participants also reported how education and experience gave them the knowledge and skills to analyze qualitative reactions. For example, Participant G shared how her undergraduate degree program had a strong emphasis on data analysis, and how she had "always enjoyed doing analyses," while Participant H shared that he had obtained a certification from the ROI Institute, Inc. and had "a passion for this area." Participant J explained how previous experience played a part in what she had done throughout her career:

> Well, I've learned from all the other companies that I have belonged to—like [previous employer]—I told you just really gave me a great foundation. We had a great training program there, so I've borrowed a lot of the things that we've done with them, and I've kind of incorporated it into every job that I've been to.

Support systems. Most participants reported having some type of support system in place to help with the data analysis. This support fell into two categories. First, some participants received support from individuals inside their teams, such as Participant B, who said that "it's really the entire team's responsibility… we kind of all have a role within evaluations and reporting, whether that's actually like building out the evaluation, or delivering it, or analyzing it, everyone kind of has a hand to play in it." Participant G shared that her team "decide[s] we're going to review everything or we'll review a certain aspect of something, so that's nice because it's definitely a team effort and it's helpful to work with other people." Second, some participants received support from individuals outside their teams, such as Participant I, who explained that although she analyzed the data set, she was also required to share it with her organization's headquarters so they could enter it into a tracking system and use it for performance management purposes:

> Anything that a student writes in those open-ended questions, they have to enter it, even if it's like an N/A, it's always entered. So, they will enter it, and then they'll… have a scoring system, and you have to be within a certain percentage type of deal because the students evaluate you, so you have to be at a certain percentage before there's continued development.

Participant E shared how he worked with stakeholders from several departments such as sales, marketing, analytics, and project management to "review [the data set] as a team and decide if we need to modify our plans going forward."

Summary

This chapter provided a summary of the results of interview responses from 13 participants. While the responses to the interview questions for RQ1 and RQ2 were

mixed, several themes emerged. For RQ1, the participants reported that they used surveying platforms to collect the data; searched for keywords, patterns, and themes in the data; and worked as a team to perform their analysis. For RQ2, the participants reported that they had learner-centric, trainer-centric, and organization-centric motives for analyzing the data; had various competencies that enabled them to perform the data analysis; and had support systems to help with the data analysis. Chapter 5 provides an interpretation of the findings and recommendations for future research.

Chapter 5

Interpretation and Recommendations

This chapter expands on the findings reported in Chapter 4. The chapter begins with a summary of the study. Next, the study's findings are discussed in relation to the literature reviewed in Chapter 2. Lastly, conclusions are presented.

Study Summary

This section provides a summary of the study. The section begins with an overview of the problem and is followed by the purpose statement and research questions. Next, the methodology used to carry out the study is reviewed. Finally, the section concludes with an overview of the major findings.

Overview of the problem. Although Level 1 evaluation is the most commonly used training evaluation method (Association for Talent Development, 2016; Brown, 2005; Chapman, 2000; Gomez, 2003; Hypes, 2016; Moller & Mallin; 1996; Phillips, 2000), there is a relative lack of research on Level 1 evaluation itself, including the popular reactionnaire, which is routinely used to conduct Level 1 evaluation. Although some literature does exist on Level 1 evaluation, it largely centers on the quantitative data captured via reactionnaires and stops short in exploring any qualitative data collected from the instrument. Nevertheless, qualitative reactions can provide some of the most critical training program evaluation information (J.L. Pershing, 2006; Kirkpatrick, 2008; Kirkpatrick & Kirkpatrick, 2009; Thalheimer, 2016c). Despite the usefulness of qualitative reactions, little attention has been paid to the analysis of this data. A review of the literature reveals a set of rudimentary guidelines that instructional designers can

follow when analyzing the data; however, no literature or research appears to exist on how instructional designers analyze qualitative reactions at present.

Purpose statement and research questions. There were two purposes for this study. The first purpose was to examine the methods used by instructional designers to analyze qualitative reactions. The second purpose was to examine the underlying reasons for the use of the methods of analysis. Two research questions guided the study:

RQ1: How do instructional designers analyze qualitative reactions?

RQ2: What rationale do instructional designers provide for the methods they use to analyze qualitative reactions?

Review of the methodology. This study was approached from a pragmatic standpoint which positioned the researcher as a human performance improvement practitioner seeking to uncover the participants' current state of performance related to the phenomenon. A conceptual framework that included a means to describe the phenomenon was developed (see Figure 6) and used to create a semi-structured interview protocol with 11 questions. Thirteen participants were purposively selected for interviews. The interview data was coded and analyzed in MAXQDA using the same conceptual framework developed to describe the phenomenon. While the research was carried out, steps were taken to ensure the reliability and trustworthiness of the study.

Major findings. For RQ1, participants shared various processes and methods to analyze qualitative reactions, and three themes emerged. First, most participants used at least one surveying platform such as Survey Monkey, Microsoft Forms, or Qualtrics to collect the qualitative reactions. Second, most participants used a range of tactics to look for keywords, patterns, or themes in the data. Third, most participants reviewed the data

as a team because it either helped them validate any individual analyses or provided a means by which they could gather more context to better understand and interpret the data.

For RQ2, participants shared various underlying reasons for using the QDA processes they described, and three themes emerged. First, all participants had one or more primary motives for analyzing the data. These motives fell into three categories:

- Learner-centric motives, which were tied to the learners themselves.
- Trainer-centric motives, which were tied to the trainers themselves, their performance, and their growth and development.
- Organization-centric motives, which were tied to the participants' departmental or organizational needs and goals.

Second, most participants incorporated their unique skills, knowledge, and abilities when analyzing their data sets. Third, most participants had a support system to help with the data analysis; participants received support from individuals inside or outside their direct teams.

Findings Related to the Literature

A review of the literature was conducted to examine the methods instructional designers use to analyze qualitative reactions and the reasons for using those methods. The literature review exposed a lack of research on the topic, which prompted this exploratory, descriptive study. The following sections describe how the themes that emerged from the interviews are related to the literature reviewed in Chapter 2.

Emerging themes from Research Question 1. Coffey and Atkinson (1996), Miles et al. (2019), and Patton (2015) emphasized that there is no correct way to perform

QDA and that one must choose the analysis method that best fits the use case. The results of this study support that assertion. Figure 8 displays the emerging themes from RQ1 in relation to the theoretical framework used to conduct the performance analysis (see Figure 6 for the theoretical framework). Although the participants similarly used surveying platforms to compile the data, they used a variety of processes and methods to make sense of the data. While only one participant specifically used the term *coding*, most participants performed coding activities that helped them disassemble and reassemble the data to discover patterns and themes. This finding reinforces the supposition that coding is a fundamental activity in QDA, as indicated by Auerbach and Silverstein (2003), Creswell and Creswell (2018), Grbich (2007), Miles et al. (2019), Saldaña (2015), and Tesch (2013).

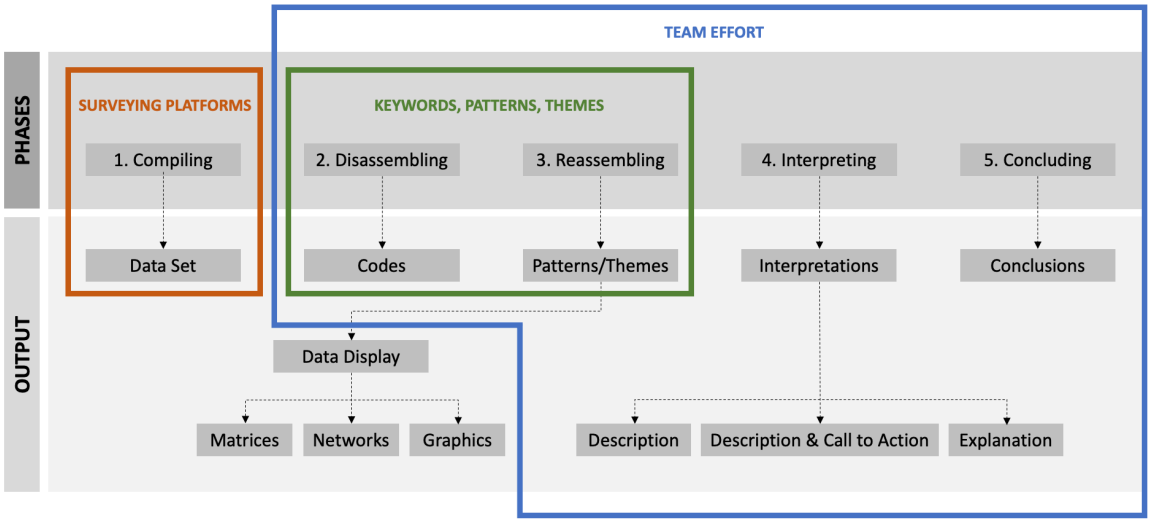


Figure 8. RQ1 themes and their relation to the theoretical framework for performance analysis. Note that some connecting lines have been removed from the framework to increase the readability of the figure.

Yin (2015) offered three broad types of interpretation: *description*, *description plus a call for action*, and *explanation*. The results of this study demonstrate that these categories of interpretation align with the interpretations instructional designers draw from the analysis of qualitative reactions. The participants described interpretive actions that fall neatly under each category and described various ways in which their interpretations and conclusions were reported to stakeholders. Unsurprisingly, when the participants presented their findings to stakeholders, most presented the results of their qualitative analysis in narrative format, supporting Kirkpatrick and Kirkpatrick's (2015b) claim that when reporting data, one might include trainee quotes to illuminate the quantitative data compiled from reactionnaires. Conversely, two participants heeded Newby's (1992) directive to determine the frequencies of categorized qualitative reactions, though the participants' quantification of this data stemmed from the expectations and processes established by their respective organizations and not from Newby's guidance.

The team effort that occurred during the analysis of the qualitative reactions was unexpected. Figure 9 displays the relationship between the two primary reasons reported for the team effort and the theoretical framework created to describe the phenomenon (see Figure 6 for the conceptual framework). During the disassembling and reassembling phases, the participants relied on their teammates to verify the accuracy of their analyses, fundamentally using peer debriefing (Creswell & Creswell, 2018; Lincoln & Guba, 1985) to help guarantee their analyses were reliable and trustworthy. During the interpreting and concluding phases, the participants relied on peers to provide more context to help

them better understand and interpret the data. By providing this context, the peers assisted in producing one of Yin's (2015) interpretations.

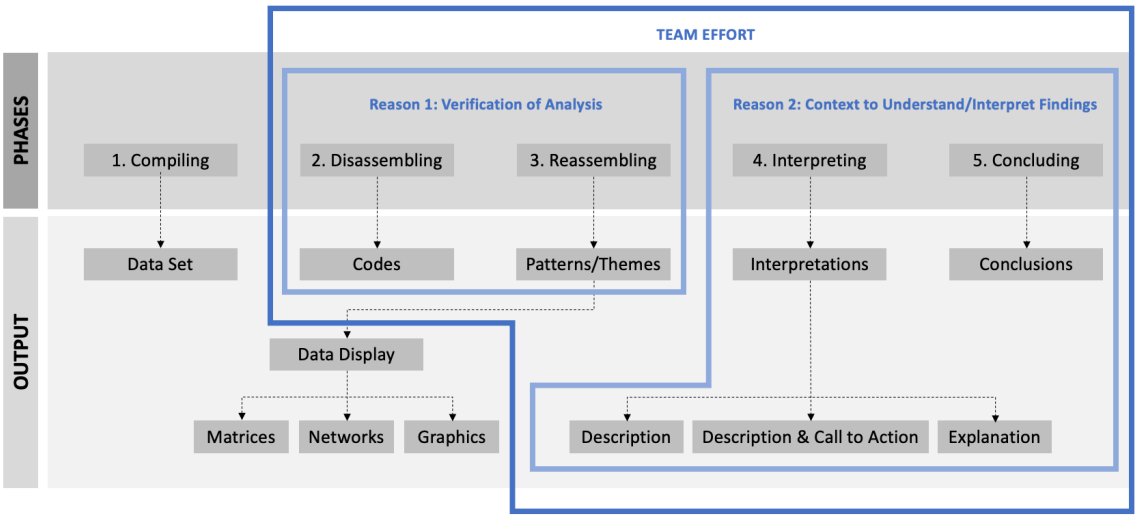


Figure 9. The relationship between the two primary reasons reported for team effort and the theoretical framework for performance analysis. Note that some connecting lines have been removed from the framework to increase the readability of the figure.

Emerging themes from Research Question 2. Figure 10 displays the emerging themes from RQ2 in relation to the conceptual framework used to conduct the cause analysis (see Figure 6 for the conceptual framework). Chevalier (2003) emphasized that environmental factors "pose the greatest barriers to exemplary performance" (p. 4). The results of this study suggest a general lack of barriers, as evidenced by the participants' focus on the helpful support systems they had in place. In contrast, the individual factors appear to play a more prominent role in how the participants analyzed the qualitative reactions.

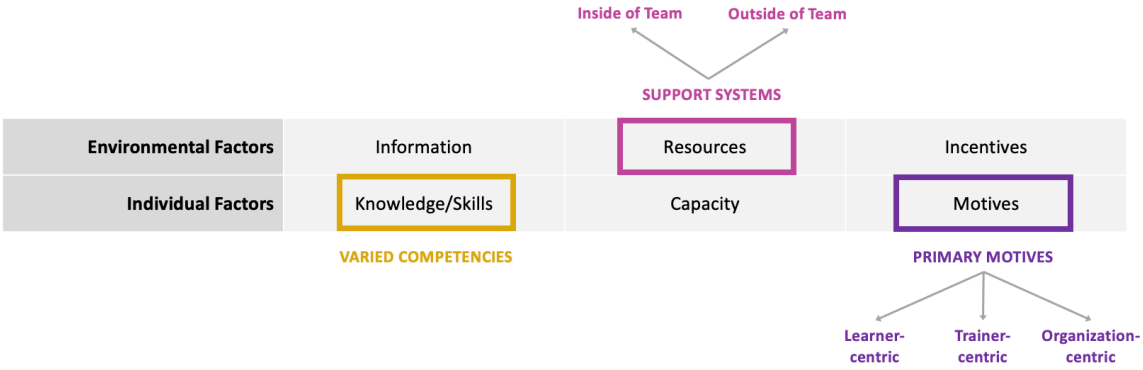


Figure 10. RQ2 themes and their relation to the conceptual framework for cause analysis.

A review of the literature on Level 1 evaluation revealed that it is useful for three reasons. First, reactions help determine learner satisfaction (Brinkerhoff, as cited in Thalheimer, 2016a; Thalheimer, 2016b; Kirkpatrick, 2006; Phillips & Phillips, 2010; Phillips & Phillips, 2016). Second, reactions can inform the design and improvement of training programs (Brinkerhoff, as cited in Thalheimer, 2016a; Thalheimer, 2016b; Kirkpatrick & Kirkpatrick, 2015b; Kraiger, 2002; Morgan & Casper, 2000; Pershing & Pershing, 2001; Phillips & Phillips, 2010; Phillips & Phillips, 2016). Third, reactions are useful in substantiating a program's justification, value, or worth (Kirkpatrick & Kirkpatrick, 2015b; Morgan & Casper, 2000; Phillips & Phillips, 2010; Phillips & Phillips, 2016; Thalheimer, 2016b). The results of this study provide evidence of a link between the usefulness of Level 1 evaluation and the three primary categories of motives for analyzing qualitative reactions, as displayed in Figure 11. The participants' learner-centric motives align with the usefulness of Level 1 evaluation in both determining learner satisfaction and informing the design and improvement of training, while their

trainer-centric motives align with informing the design and improvement of training and their organization-centric motives align with substantiating the justification, value, or worth of training.

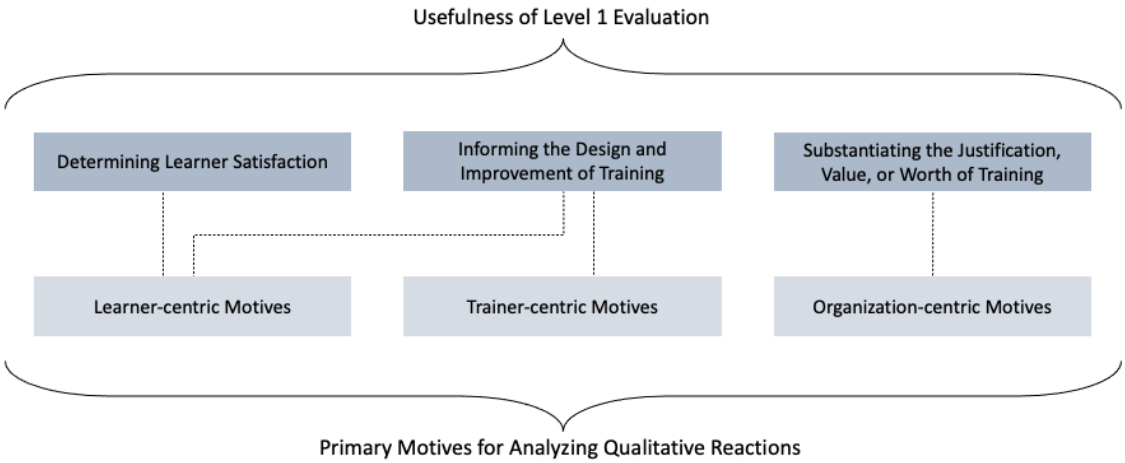


Figure 11. Link between the usefulness of Level 1 evaluation and the primary motives for analyzing qualitative reactions.

Other emerging themes. Although not directly tied to the methods used to analyze qualitative reactions or the rationale for using those methods, two additional themes emerged from the interviews: challenges with learner feedback and non-mandatory reporting requirements.

Challenges with learner feedback. Most participants indicated they experienced challenges in obtaining qualitative reactions, understanding qualitative reactions, or both. For example, Participant D explained how she experienced challenges not only with "getting the feedback" but also "translating it" because "some people aren't so great at writing anymore." Participant L explained how it could also be "something as simple

as—like if they're filling them out by hand—handwriting. You know, actually being able to read them." Participant L also described additional challenges such as student motivation to provide feedback and the clarity of that feedback:

> Another thing would be students not taking it seriously, even though we preface it like, "Hey, you guys, we do look at these; they are evaluated and reviewed." Or maybe students not being that specific. They may give some sort of developmental feedback like the course was too long. So, what exactly was too long? Was it the length of the day? Was it the length of the course? Would more breaks have been beneficial? More self-study? It's just they'll give a nugget, but we don't have an opportunity to follow up to get that information from them.

Some participants also mentioned how they took any feedback with a grain of salt, such as Participant M:

> You always have someone who doesn't like what you're doing or why you're doing it. And those tend to be the most vociferous of the respondents and open-ended survey questions. And, you know, there's a sense of pride that goes into our work, of course, and the challenge in that is being objective and not taking it personally, but also taking it with a grain of salt, because we know this person has a habit of ripping things up and handing them back because they don't believe training is important. They don't believe that what we do is worthwhile.

Like Participant M, Participant E also shared how remaining objective can be challenging:

> I think the challenge is to try to read them and understand them using good communication principles, like understand what they're trying to say. I think it's

> really easy when reading an open-ended question and a survey to project what you want or think the situation to be and use it to validate some preconceived notions, so trying to keep that more—it's almost like the scientific mentality, like anything's possible here. Don't go into it with a preconceived notion; let the comments and the data in the survey tell you what the situation is.

Non-mandatory reporting requirements. Although most participants indicated they were not required to report their findings to stakeholders, they opted to share this information anyway. For example, Participant J said, "It's just something that we do… we're still a really small company, and we kind of make our own rules, which I enjoy… and our executive team really stand behind training." Participant H noted that "it's a nice to do type of thing" and elaborated on why he chose to share the information with stakeholders and learners:

> For me, I do like to go back and I like to share the results back with number one, the management team, so they know what they're putting their people through. I also go back to the learners and I also go over the results with them. So really quick, maybe about like two minutes, but… what I want to do is make sure that I thank them for their time in terms of providing the input. And then the other reason that I do that is I want to make sure that they know they're not just taking an evaluation where I'm not going to do anything with it.

Some participants who shared their findings with stakeholders or learners did so on an ad-hoc basis, whereas others shared their findings on a more consistent basis, such as monthly (Participant K) or bi-annually (Participant G). These findings were primarily presented in narrative form; none of the participants used data displays as part of the

reporting process. For example, Participant B shared how she used quotes to present the themes that emerged from her analysis: “If there is a specific comment that really supports that theme, then I’ll put it in there with quotations and provide that as well." Only two participants indicated they presented the qualitative reactions in a quantitative format. Participants C and G both used processes that involved coding the data in Excel and then performing numerical analyses using the codes. Participant G shared that with her coding process, "you can easily count the number of people who said they want to change X, and then we can calculate a percentage of employees who are looking for this or would change this." Participant C shared how quantifying the qualitative reactions can be helpful when presenting the data to stakeholders:

> Cause while they're not numbers, they sort of are, and numbers just don't lie. And it's a little subjective, but words have meaning, and they have power, and if you can kind of quantify that in a way, you will just get buy-in. There's no other way.

The relationship of the other emergent themes to RQ1 and RQ2. Although the results of the study suggest a general lack of environmental barriers, the participants did express challenges in obtaining qualitative reactions, understanding qualitative reactions, or both. While the challenges with learner feedback had no direct impact on the QDA methods the participants described, they can be indirectly associated with the environmental factor of *resources* since the data itself must be available for an analysis to take place. Additionally, the general lack of environmental barriers may be linked to the fact that the participants were not required to report their findings to training program stakeholders. In other words, it might be possible that because stakeholders did not communicate reporting requirements or expectations—which fall under the

environmental factor of *information*—the participants had the flexibility to adopt methods of analysis that were largely driven by their knowledge, skills, and motives (though some participants' methods were influenced by their manager's goals or their organization's appetite for data).

Conclusions

This study was exploratory and descriptive, shedding light on how instructional designers analyze qualitative reactions. The study's findings prompt practical implications for action and considerations for future research. This section discusses these notions in more detail and finishes with concluding remarks.

Implications for action. The results of this study present three practical implications for action centered on educating instructional designers on QDA practices. First, the study may serve as a starting point for instructional designers who do not analyze qualitative reactions. Second, the study may help instructional designers identify ways in which they can improve their QDA processes. For example, instructional designers who currently perform this activity on an individual basis may see value in incorporating other team members into the QDA process. Third, the study may guide instructional designers and professional instructional design associations in pinpointing QDA standards, professional development activities, or both. While the participants in this study described a variety of QDA methods, they did so mainly without using QDA terms commonly found in the literature, which may indicate a lack of a more formal understanding of QDA. Given the diversity of QDA approaches and strategies, instructional designers could benefit from a shared, foundational understanding of QDA processes.

Recommendations for future research. The results of this study present several opportunities for future research. First, because this study was limited to 13 participants, a future study could be expanded to include a larger number of participants. A larger sample size could more closely represent the population, which could increase the generalizability of the results. Second, because the study was conducted during the COVID-19 pandemic and limited the participants' experiences with the phenomenon to the use of electronic reactionnaires, a future study could explore how instructional designers who collect paper-based reactionnaires analyze qualitative reactions. Additionally, the results of that study could be compared to the results of the current study. Third, a future study could investigate the differences in QDA processes related to the number of qualitative reactions an instructional designer must analyze. When analyzing quantitative reactionnaire data, an instructional designer will likely calculate averages, and the process for calculating averages remains the same regardless of the number of data points. However, an abundance of qualitative reactions—such as hundreds, if not thousands, of qualitative reactions from an eLearning course—might be analyzed very differently than only a handful of qualitative reactions collected from a small training program. Fourth, a future study could investigate the differences in QDA processes related to specific instructional designer contexts. For example, future research could compare how instructional designers' QDA processes vary by industry, level of education, or work experience. Fifth, a future study might investigate the connections between quantitative reactionnaire data and qualitative reactions. Finally, a future study might explore how instructional designers use the results of their analysis of qualitative reactions or the challenges they run into with the analyzed data or both.

Concluding remarks. Although it has been more than 60 years since it was first introduced, the Kirkpatrick model remains ever popular today, and the literature confirms that Level 1 evaluation is the most widely used training evaluation method. Despite this fact, researchers have shied away from studying Level 1 evaluation and have opted to research less used levels of the Kirkpatrick model and less popular evaluation models. However, Bates (2004) and Giangreco et al. (2010) reasoned that scholars must research the types of training evaluations in use. This study contributed to the literature on Level 1 evaluation by providing insight into an under-researched aspect of that level of evaluation. While the results of the study reveal that participants used a variety of QDA processes to analyze qualitative reactions, they also imply a lack of a more formal understanding of QDA. Thus, this study serves as a starting point in further exploring the paradox surrounding the analysis of qualitative reactions.

References

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42*(2), 331–342. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1111/j.1744-6570.1989.tb00661.x

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

Association for Talent Development (2016). *Evaluating learning: Getting to measurements that matter* (Whitepaper). Retrieved from: https://www.td.org/research-reports/evaluating-learning

Auerbach, C., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York, NY: New York University Press.

Austin, J., & Garnier, L. (1998). The virtual office: A Behavior Engineering Model (BEM) perspective. *Performance Improvement Quarterly, 11*(4), 7–21. Retrieved from https://search-ebscohost-com.bakeru.idm.oclc.org/login.aspx?direct=true&db=eric&AN=EJ597493&site=ehost-live

Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation & Program Planning*, *27*(3), 341–347. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1016/j.evalprogplan.2004.04.011

Blanchard, P. N., Thacker, J. W., & Way, S. A. (2000). Training evaluation: Perspectives and evidence from Canada. *International Journal of Training & Development, 4*(4), 295. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1111/1468-2419.00115

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management, 36*(4), 1065–1105. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1177/0149206309352880

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1191/1478088706qp063oa

Brinkerhoff, R. O., & Dressler, D. (2015). Using evaluation to build organizational performance and learning capability: A strategy and a method. *Performance Improvement, 54*(7), 37–44. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/pfi.21497

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. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1037/0021-9010.90.5.991

Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching & Learning, 10*(6), 807–815. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1016/j.cptl.2018.03.019

Chapman, D. D. (2000). *Preferences of training performance measurement: A comparative study of North Carolina training professionals and non-training managers* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection (Order No. 9992460). Retrieved from https://bakeru.idm.oclc.org/login?url=https://search-proquest-com.bakeru.idm.oclc.org/docview/304610918?accountid=26368

Chen, H. L. (2018). Evaluation versus research. In B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 630–632). San Francisco, CA; SAGE Publications Inc.

Chen, K.-Z., Lowenthal, P. R., Bauer, C., Heaps, A., & Nielsen, C. (2017). Moving beyond smile sheets: A case study on the evaluation and iterative improvement of an online faculty development program. *Online Learning*, *21*(1), 85–111. Retrieved from https://bakeru.idm.oclc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1140246&site=ehost-live

Chevalier, R. (2003). Updating the Behavior Engineering Model. *Performance Improvement, 42(5*), 1–16. Retrieved from https://www.researchgate.net/publication/229651267_Updating_the_behavior_engineering_model

Clegg, W. H. (1987). Management training evaluation: An update. *Training & Development Journal, 41*(2), 65–71. Retrieved from https://search-ebscohost-com.bakeru.idm.oclc.org/login.aspx?direct=true&db=aph&AN=9127210&site=ehost-live

Clement, R. W. (1982). Testing the hierarchy theory of training evaluation: An expanded role for trainee reactions. *Public Personnel Management, 11*(2), 176–184. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1177/009102608201100210

Coffey, A., & Atkinson, P. (1996) *Making sense of qualitative data analysis: Complementary strategies*. Thousand Oaks, CA: Sage Publications, Inc.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed). Thousand Oaks, CA: Sage Publications, Inc.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage Publications, Inc.

Dessinger, J. C., & Moseley, J. L (2006). The full scoop on full-scope evaluation. In J.A. Pershing (Ed.), *Handbook of human performance technology: Principles, practices, and potential* (pp. 312–330). San Francisco, CA: Pfeiffer.

DeVaughn, P., & Stefaniak, J. (2020). An exploration of the challenges instructional designers encounter while conducting evaluations. *Performance Improvement Quarterly*. Retrieved from https://doi.org/10.1002/piq.21332

Eseryel, D. (2002). Approaches to evaluation of training: Theory & practice. *Journal of Educational Technology & Society, 5*(2), 93–98. Retrieved from https://www.semanticscholar.org/paper/Approaches-to-Evaluation-of-Training%3A-Theory-%26-Eseryel/0a4444e08e3891f4d9e2d27f584bb942d9a7a0a4

Foreman, S. M. (2008). *Kirkpatrick model: Training evaluation practices in the pharmaceutical industry* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection (Order No. 3304485). Retrieved from https://bakeru.idm.oclc.org/login?url=https://search-proquest-com.bakeru.idm.oclc.org/docview/304820021?accountid=26368

Furst-Bowe, J. A. (1996). An analysis of the competencies needed by trainers to use computer-based technologies and distance learning systems. *Performance Improvement Quarterly 9*(4), pp. 57–78. Retrieved from https://web-a-ebscohost-com.bakeru.idm.oclc.org/ehost/detail/detail?vid=6&sid=d76b2073-40da-4744-8f88-df9dd510b88c%40sessionmgr4008&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=EJ534500&db=eric

Giangreco, A., Sebastiano, A., & Peccei, R. (2009). Trainees' reactions to training: An analysis of the factors affecting overall satisfaction with training. *International Journal of Human Resource Management, 20*(1), 96–111. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1080/09585190802528417

Giangreco, A., Carugati, A., & Sebastiano, A. (2010). Are we doing the right thing? Food for thought on training evaluation and its context. *Personnel Review, 39*(2), 162–177. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1108/00483481011017390

Gilbert, T. F. (2007). *Human competence: Engineering worthy performance*. San Francisco, CA: Pfeiffer.

Glerum, D. R., Joseph, D. L., McKenny, A. F., & Fritzsche, B. A. (2020). The trainer matters: Cross-classified models of trainee reactions. *Journal of Applied Psychology* [Supplemental material]. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1037/apl0000503.supp

Gomez, A. K. (2003). *An analysis of the evaluation practices of employer-sponsored training in the financial services industry* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection (Order No. 3096741). Retrieved from https://bakeru.idm.oclc.org/login?url=https://search-proquest-com.bakeru.idm.oclc.org/docview/305245962?accountid=26368

Grbich, C. (2007). *Qualitative data analysis: An introduction*. Thousand Oaks, CA: Sage Publications, Inc.

Guerra-López, I. (2008). *Performance evaluation: Proven approaches for improving program and organizational performance*. San Francisco, CA: Jossey-Bass.

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. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1037/a0038380

Hart, T., & Achterman, P. (2017). Qualitative Analysis Software. In J. Matthes (Ed.), *The international encyclopedia of communication research methods*. Hoboken, NJ: John Wiley & Sons, Inc.

Hellebrandt, J., & Russell, J. D. (1993). Confirmative evaluation of instructional materials and learners. *Performance and Instruction, 32*(6), 22–27. Retrieved from https://doi.org/10.1002/pfi.4170320608

Henderson, S., & Segal, E. H. (2013). Visualizing qualitative data in evaluation research. *New Directions for Evaluation, 2013*(139), 53–71. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/ev.20067

Holton III, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, *7*(1), 5–21. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/hrdq.3920070103

Hypes, B. R. (2016). *Training programs in the federal government: A systematic review of how effectiveness is evaluated* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection (Order No. 10252416). Retrieved from https://bakeru.idm.oclc.org/login?url=https://search-proquest-com.bakeru.idm.oclc.org/docview/1884648930?accountid=26368

International Association for Continuing Education and Training (IACET) (2018). *ANSI/IACET 1-2018 standard for continuing education and training*. Retrieved from https://www.iacet.org

International Board of Standards for Training & Performance Improvement (IBSTPI) (2013). *2012 IBSTPI instructional designer standards: Competencies & performance statements*. Retrieved from https://ibstpi.org

Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: Beyond Kirkpatrick. *Human Resource Development Quarterly*, *5*(4), 371–380. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/hrdq.3920050408

Kennedy, P. E., Chyung, S. Y., Winiecki, D. J., & Brinkerhoff, R. O. (2014). Training professionals' usage and understanding of Kirkpatrick's Level 3 and Level 4 evaluations. *International Journal of Training & Development, 18*(1), 1–21. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1111/ijtd.12023

Kim, S., Park, S., Lavelle, J., Kim, M., & Chaudhuri, S. (2020). Revisiting trainee reactions: A multilevel analysis of the nomological network. H*uman Resource Development Quarterly, 31*(2), 173–191. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/hrdq.21383

King, J.A., Stevahn, L., Ghere, G., & Minnema, J. (2001). Toward a taxonomy of essential evaluator competencies. *American Journal of Evaluation, 22*(2), 229–247. Retrieved from https://bumoodle.bakeru.edu/mod/folder/view.php?id=667494

Kirkpatrick, D. L. (2006). Seven keys to unlock the four levels of evaluation. *Performance Improvement, 45*(7), 5–8. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/pfi.2006.4930450702

Kirkpatrick, J. D. (2008). *The new world level 1 reaction sheets.* Kirkpatrick Partners. Retrieved from https://www.kirkpatrickpartners.com/Portals/0/Resources/Library/The%20New%20World%20Level%201%20Reaction%20Sheets.pdf

Kirkpatrick, D. L., & Kirkpatrick, J.D. (2006). *Evaluating training programs: The four levels* (3rd ed.). San Francisco, CA: Berrett-Koehler Publications, Inc.

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2009). *Kirkpatrick then and now: A strong foundation for the future.* St. Louis, MO: Kirkpatrick Partners, LLC.

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2015a). *Build a better reaction sheet: Tips from Jim and Wendy Kirkpatrick* [Webinar]. Training Magazine Network. Retrieved from https://www.trainingmagnetwork.com/events/423

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2015b). The four levels of evaluation – an update. *TD at Work, 32*(1502). Retrieved from https://www.td.org/td-work/the-four-levels-of-evaluation-an-update

Klein, J. D., & Kelly, W. Q. (2018). Competencies for instructional designers: A view from employers. *Performance Improvement Quarterly, 31*(3), 225–247. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/piq.21257

Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice* (pp. 331–376). San Francisco: Jossey-Bass.

Lee, S. H. (1998). *Making reaction evaluation a more useful tool in the evaluation of corporate training programs: Reactionnaire dimensions and design criteria* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection (Order No. 9907312). Retrieved from https://bakeru.idm.oclc.org/login?url=https://search-proquest-com.bakeru.idm.oclc.org/docview/304413761?accountid=26368

Lee, S. H. (2006). Constructing effective questionnaires. In J.A. Pershing (Ed.), *Handbook of human performance technology: Principles, practices, and potential* (pp. 760–779). San Francisco, CA: Pfeiffer.

Lee, S. H., & Pershing, J. *A*. (2002). Dimensions and design criteria for developing training reaction evaluations. *Human Resource Development International, 5*(2), 175–197. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1080/13678860110053678

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications, Inc.

Long, L. N. (1990). *A test of Kirkpatrick's training evaluation levels in terms of utility for decision-making* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection (Order No. 9030722). Retrieved from https://bakeru.idm.oclc.org/login?url=https://search-proquest-com.bakeru.idm.oclc.org/docview/303875527?accountid=26368

Lunenburg, F. C., & Irby, B. J. (2008). *Writing a successful thesis or dissertation: Tips and strategies for students in the social and behavioral sciences*. Thousand Oaks, CA: Corwin Press.

Maxwell, J. A. (2018). Qualitative data analysis. In B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1335–1339). San Francisco, CA; SAGE Publications Inc.

McLinden, D., & Boone, W. J. (2009). More than smile sheets: Rasch analysis of training reactions in a medical center. *Performance Improvement Quarterly*, *22*(3), 7–21. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/piq.20042

Miles, M. B., Huberman, A. M., & Saldaña, J. (2019) *Qualitative data analysis: A methods sourcebook* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Misanchuk, E. (1978). Descriptors of evaluations in instructional development: Beyond the formative-summative distinction. *Journal of Instructional Development, 2*(1), 15–19. Retrieved from http://www.jstor.org/stable/30220544

Moller, L., & Mallin, P. (1996). Evaluation practices of instructional designers and organizational supports and barriers. *Performance Improvement Quarterly, 9*(4), 82–92. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1111/j.1937-8327.1996.tb00740.x

Morgan, R. B., & Casper, W. J. (2000). Examining the factor structure of participant reactions to training: A multidimensional approach. *Human Resource Development Quarterly, 11*(3), 301–317. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/1532-1096(200023)11:3<301::AID-HRDQ7>3.0.CO;2-P

Newby, A. C. (1992). *Training evaluation handbook.* Aldershot, England: Gower Publishing Company Limited.

Newstrom, J. W. (1978). Catch-22: The problems of incomplete evaluation of training. *Training & Development Journal, 32*(11), 22–24. Retrieved from http://search.ebscohost.com.bakeru.idm.oclc.org/login.aspx?direct=true&db=asn&AN=9068707&site=ehost-live

Patton, M.Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Pershing, J. A. (2006). Human performance technology fundamentals. In J.A. Pershing (Ed.), *Handbook of human performance technology: Principles, practices, and potential* (pp. 5–92). San Francisco, CA: Pfeiffer.

Pershing, J. L. (2006). Performance measurement and assessment. In J.A. Pershing (Ed.), *Handbook of human performance technology: Principles, practices, and potential* (pp. 745–759). San Francisco, CA: Pfeiffer.

Pershing, J. A., & Pershing, J. L. (2001). Ineffective reaction evaluation. *Human Resource Development Quarterly, 12*(1), 73–90. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1002/1532-1096(200101/02)12:1<73::AID-HRDQ6>3.0.CO;2-7

Phillips, J. H. (2000). *Evaluating training programs for organizational impact: Five reports* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection (Order No. 9966169). Retrieved from https://bakeru.idm.oclc.org/login?url=https://search-proquest-com.bakeru.idm.oclc.org/docview/304631328?accountid=26368

Phillips, J. J., & Phillips, P. P. (2010). *Measuring for success: What CEOs really think about learning investments.* Alexandria, VA: ATD Press.

Phillips, J. J., & Phillips, P. P. (2016). *Handbook of training evaluation and measurement methods* (4th ed.). Abingdon, England: Routledge.

Pulichino, J. P. (2007). *Usage and value of Kirkpatrick's four levels of training evaluation* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global: The Humanities and Social Sciences Collection (Order No.

3264775). Retrieved from https://bakeru.idm.oclc.org/login?url=https://search-proquest-com.bakeru.idm.oclc.org/docview/304707180?accountid=26368

Reiser, R. A. (2018). What field did you say you were in? In R. A. Reiser, & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology* (4th ed., pp. 1–7). London, UK: Pearson.

Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance and change.* New York, NY: Basic Books.

Saldaña, J. (2015). *The coding manual for qualitative researchers (*3rd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Scriven, M. (1966, March). Social Science Education Consortium. Publication 110, *The Methodology of Evaluation*. Retrieved from https://eric.ed.gov/?id=ED014001

Slone, D. J. (2009). Visualizing qualitative information. *Qualitative Report, 14*(3), 489–497. Retrieved from http://search.ebscohost.com.bakeru.idm.oclc.org/login.aspx?direct=true&db=asn&AN=45032156&site=ehost-live

Silverman, D. (2015). *Interpreting qualitative data.* Thousand Oaks, CA: Sage Publications, Inc.

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. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1037/0021-9010.93.2.280

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2nd ed.). San Francisco, CA: Jossey-Bass.

Tan, J. A., Hall, R. J., & Boyce, C. (2003). The role of employee reactions in predicting training effectiveness. *Human Resource Development Quarterly, 14*(4), 397–411. Retrieved from https://web-b-ebscohost-com.bakeru.idm.oclc.org/ehost/detail/detail?vid=13&sid=81e5fb9a-b6a3-4dc1-9b02-410d9ab371e6%40pdc-v-sessmgr05&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=EJ772887&db=eric

Tesch, R. (2013). *Qualitative research: Analysis types and software*. Abingdon, England: Routledge.

Thalheimer, W. (2016a). *Performance-focused smile sheets: A radical rethinking of a dangerous art form.* Somerville, MA: Work-Learning Press.

Thalheimer, W. (2016b, September 28). *Smile sheet apocalypse: Why your smile sheets are so ineffective!* eLearning Industry. Retrieved from https://www.convergencetraining.com/blog/evidence-based-training-methods-and-smile-sheets-an-interview-with-dr-will-thalheimer

Thalheimer, W. (2016c). Smile sheet questions – New examples July 2016. Retrieved from https://www.worklearning.com/2016/07/06/smile-sheet-questions-new-examples-july-2016/

Thomas, M. N. (2006). Quantitative data analysis. In J.A. Pershing (Ed.), *Handbook of human performance technology: Principles, practices, and potential* (pp. 837–872). San Francisco, CA: Pfeiffer.

Verdinelli, S., & Scagnoli, N. I. (2013). Data display in qualitative research. *International Journal of Qualitative Methods, 12*(1), 359–381. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1177/160940691301200117

Warr, P., Allan, C., & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational & Organizational Psychology, 72*(3), 351–375. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1348/096317999166725

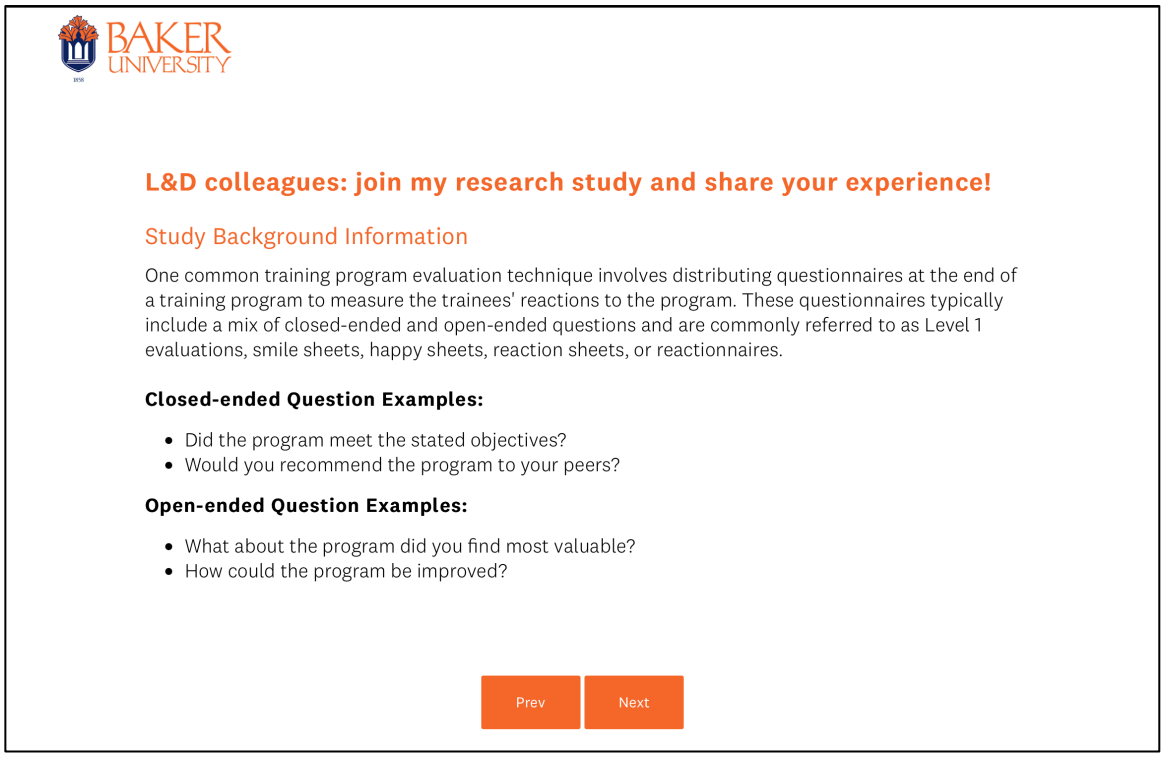
Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology, 48*(2), 347–375. Retrieved from https://doi-org.bakeru.idm.oclc.org/10.1111/j.1744-6570.1995.tb01761.x

Williamson T., & Long A.F. (2005). Qualitative data analysis using data displays. *Nurse Researcher, 12*(3), 7–19. Retrieved from https://search-proquest-com.bakeru.idm.oclc.org/docview/200779744/fulltextPDF/DA1ECBA9088340B2PQ/1?accountid=26368

Yin, R. K. (2015). *Qualitative research from start to finish* (2nd ed.). New York, NY: Guilford Press.

Appendices

Appendix A. Participant Recruitment Survey



BAKER UNIVERSITY

L&D colleagues: join my research study and share your experience!

Informed Consent Form

Purpose
The purpose of this research study is to help uncover how learning and development practitioners analyze the written comments received on training feedback forms.

Participation
Your participation in this research study is completely voluntary; you may choose not to participate. If you choose to participate, you may withdraw at any time. If you choose not to participate or if you withdraw from the study at any time, you will not be penalized.

Benefits, Risks, and Discomforts
This study adds to our understanding of how learning and development practitioners analyze qualitative reactions. There are no known risks or discomforts involved with your participation in this study. There are no direct benefits to you as a participant in this study.

Compensation
You will receive no compensation for participating in the survey portion of this study. If in the survey you indicate your interest in participating in the interview portion of the study, are selected for an interview, and complete the interview in full, you will receive a $25 Amazon gift card to compensate you for your participation.

Confidentiality
The results of this study will be used for scholarly purposes only. To help protect your confidentiality, your responses in this survey will be stored in a password-protected electronic format that is only accessible to the research team. The research team includes the researcher, his major advisor, and his research analyst; your data will not be shared outside this team. Your identifying information will be stored separately from any data you provide. If you indicate you are interested in participating in an interview and then are selected for an interview, you will be asked to review a separate informed consent form and provide consent before being interviewed. Once the study concludes, your survey responses will be permanently deleted.

Questions About the Study
If you have any questions about this research study, you may contact the researcher or the researcher's major advisor.

Researcher:
Stephen Naso, Ed.D. Candidate
School of Education, Baker University
stephenjnaso@stu.bakeru.edu

Major Advisor:
Wendy Gentry, Ph.D.
School of Education, Baker University
wendy.gentry@bakeru.edu

Note: this study has been approved by Baker University's Institutional Research Board (IRB) for research involving human subjects.

Consent
Selecting the *I agree to participate* option below indicates that you have read this informed consent form, that you understand it, and that you choose to participate in the survey portion of this study.

Selecting the *I do not agree to participate* option below indicates that you do not wish to participate in the survey portion of this study.

1. Please indicate your consent for this study.

- ◯ I agree to participate.
- ◯ I do not agree to participate.

Prev Next

BAKER UNIVERSITY

L&D colleagues: join my research study and share your experience!

Interviewee Questions

2. In what state or U.S. territory do you live?

3. What is your official job title at your organization?

4. Does your job involve the design, development, implementation, evaluation, and/or management of training?

- ◯ Yes
- ◯ No
- ◯ Other (please specify)

5. Which of the following best describes the type of organization you currently work for?

6. Which of the following best describes the principal industry of your organization?

7. Training programs are formal learning events that are goal-oriented, planned, and structured. In the past 6 months, did your organization offer your learners any training programs? If **No**, you can skip questions 8 and 9.

- ◯ Yes
- ◯ No

8. Approximately how many training programs did you offer in the past 6 months?

9. Did you evaluate any of these training programs using questionnaires that included open-ended questions?

- ◯ Yes
- ◯ No

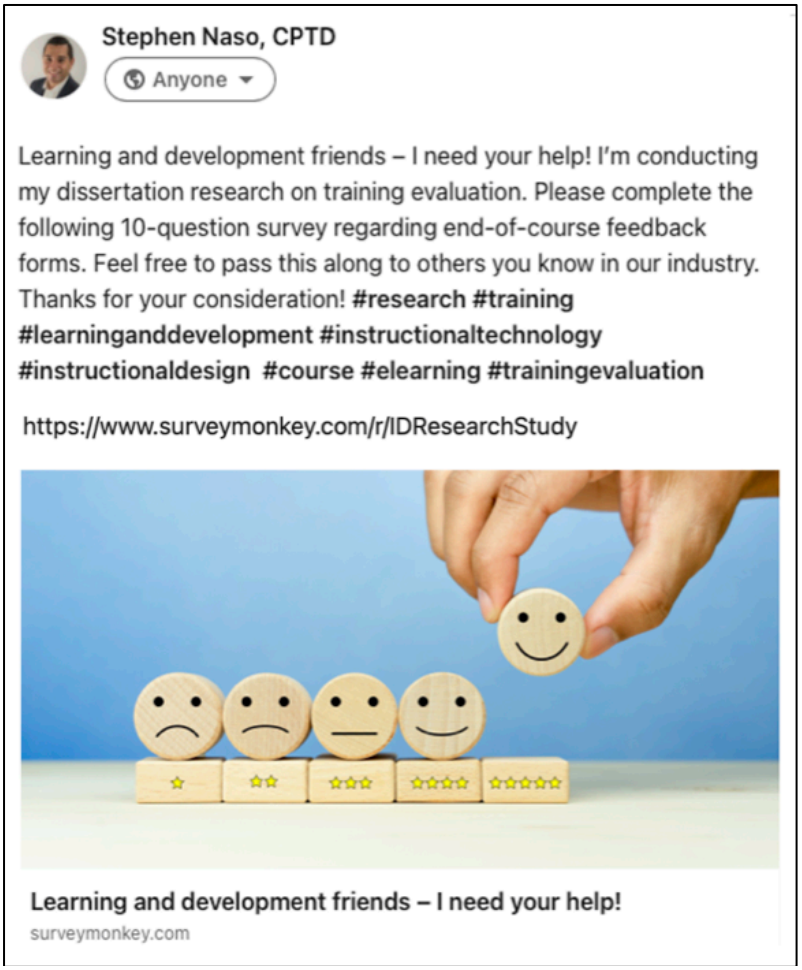
10. Are you interested in participating in a voluntary interview of up to 60 minutes with the researcher? If so, please enter your name and email address. If you are selected for an interview, the researcher will contact you via email and will provide additional information about the interview process. Once you complete the interview, you will receive a $25 Amazon gift card to compensate you for your participation.

Name

Email Address

Prev Done

Appendix B. Social Media Post for Participant Recruitment



Appendix C. Interview Protocol

Interview Protocol

Research questions:

RQ1: How do instructional designers analyze qualitative reactions?
RQ2. What rationale do instructional designers provide for the methods they use to analyze qualitative reactions?

Creswell and Poth's (2016) Interview Guidelines:

- Stay within the study boundaries.
- Use the protocol to guide the interview.
- Complete the interview on time.
- Be respectful and courteous.
- Actively listen (limit questions and advice).

Opening script:

[Begin by exchanging pleasantries.]

Thank you again for participating in this study. As a reminder, the purpose of this study is to explore how individuals like you analyze answers to open-ended reactionnaire questions, and our interview should take no more than 60 minutes. Throughout the interview, is it ok if I use the term reactionnaires, or is there another term you prefer?

Before we dive into the interview questions, I'd like to take a moment to cover a few things. First, I wanted to let you know that I received your signed consent form, so thank you very much for that. Second, I wanted to let you know that if at any point you would like me to stop the interview or if you would prefer not to answer certain questions, just let me know. Third, I will be recording this interview, and I want to reassure you that this interview and your identity will remain confidential. I will also permanently delete the recording once the study is done.

Based on all of that information I shared, I have two questions for you:

1. What questions can I answer for you at this point?
2. Are you ok with me recording this interview?

[RENAME PARTICIPANT'S NAME IN ZOOM, if applicable]

Great, thanks. I just started the recording. So that we have this on record, can you confirm verbally that I have your permission to record you?

Ok, thank you. Let's get started.

Interview Questions:

Question (Prompt)	Probes and Follow-up Questions	Rationale
DQ1: Would you mind sharing a bit about the organization you work for and your role at that organization? *Note:* Job title, org type, and industry are captured in the recruitment survey.	*Probe for:* • Organization name • Organization size • Organization location • Time at organization • Time in current role • Key responsibilities	• Demographics • Establish rapport
DQ2: Would you mind sharing a bit more about your background in the learning and development field and your background evaluating training?	*Probe for:* • Years of experience in field • Years of experience evaluating training • Education (highest level of school/highest degree earned) • Interviewee location • Age • Gender	• Demographics • Establish rapport
DQ3: Tell me about the training programs you offered over the past 6 months.	*Probe for:* • Number/types of programs • Delivery type (e.g., synchronous in-person, asynchronous virtual, etc.) • Delivery schedule (i.e., frequency) • Intended audience and average participation	• Demographics • Establish rapport
DQ4: Tell me about the open-ended questions you included in your reactionnaires.	*Probe for:* • The open-ended questions that were asked • Importance of the open-ended questions (how important and why) • Amount of open-ended responses typically collected • How reactionnaires were delivered (and how responses collected)	• Demographics • Establish rapport

IQ1: After each training program ended and you collected the reactionnaires, walk me through the steps you took to analyze the open-ended responses.	*Probe for:* • Any technology used • Any reporting requirements at organization *Probing questions:* • You mentioned ____, what does that entail? • Can you explain ____ a bit further? • How do you go about making/doing ____? • Tell me what drives your decision to ____. • Tell me about a time when…	To answer RQ1.
IQ2: Tell me a bit about why you analyze open-ended responses the way you do. **IQ3:** What challenges do you face when analyzing open-ended responses? **IQ4:** What support do you receive when analyzing open-ended responses? **IQ5:** What do you think your personal strengths are when analyzing open-ended responses? **IQ6:** What motivates you to perform the analysis? **IQ7:** What do you think would help you perform your analysis better?	• What allows you to ____? • What prevents you from ____? • Tell me what drives your ability to ____. • Tell me what drives your decision to ____. • Tell me about a time when…	To answer RQ2.

Closing Script:

At this point I'm finished asking you questions. Thank you so much for your participation! The study wouldn't be possible without you.

Do you have any questions for me at this point?

Thank you again for your time today and for your willingness to participate!

Appendix D. Baker University IRB Request

BAKER UNIVERSITY

IRB Request

Date March 5, 2020 IRB Protocol Number _____ (IRB use only)

I. Research Investigator(s) (students must list faculty sponsor)

Department(s) School of Education (IDPT)

	Name	Signature	
1.	Stephen Naso		Principal Investigator
2.	Dr. Wendy Gentry		☑ Check if faculty sponsor
3.	Dr. Kyungwha Cho		☑ Check if faculty sponsor
4.			☐ Check if faculty sponsor

Principal investigator contact information

Phone 617-605-5905

Email stephenjnaso@stu.bakeru.edu

Address 12 Gates Street #1

Boston, MA 02127

Note: When submitting your finalized, signed form to the IRB, please ensure that you cc all investigators and faculty sponsors using their official Baker University (or respective organization's) email addresses.

Faculty sponsor contact information

Phone [redacted]

Email [redacted]

Expected Category of Review: ☐ Exempt ☑ Expedited ☐ Full ☐ Renewal

II. Protocol Title

Analyzing Qualitative Trainee Reactions: A Phenomenological Study of Instructional Designer Practices

III. Summary:

The following questions must be answered. Be specific about exactly what participants will experience and about the protections that have been included to safeguard participants from harm.

A. In a sentence or two, please describe the background and purpose of the research.

The purpose of this study is to examine the methods instructional designers use to analyze the qualitative data they obtain via surveys that are commonly distributed at the end of training programs. Additionally, the study examines the underlying reasons that drive the use of those methods of analysis.

B. Briefly describe each condition, manipulation, or archival data set to be included within the study.

This is a phenomenological study. Given a lack of existing research on the topic, the study is exploratory and descriptive in nature and does not include any conditions, manipulations, or archival data sets.

IV. Protocol Details

A. What measures or observations will be taken in the study? If any questionnaire or other instruments are used, provide a brief description and attach a copy.

To obtain partcipants, the researcher will use a recruitment survey that will be shared on LinkedIn (see attachment). Eligible participants will be invited to participate in semi-structured interviews. An interview protocol was developed to guide the interviews (see attachment for interview protocol).

B. Will the subjects encounter the risk of psychological, social, physical, or legal risk? If so, please describe the nature of the risk and any measures designed to mitigate that risk.

The subjects will not encouter psychological, social, physical, or legal risks.

C. Will any stress to subjects be involved? If so, please describe.

The subjects will not encounter any stress.

D. Will the subjects be deceived or misled in any way? If so, include an outline or script of the debriefing.

The subjects will not be deceived or misled in any way.

E. Will there be a request for information which subjects might consider to be personal or sensitive? If so, please include a description.

Interviewees will be asked to provide demographic information including age and gender. Prior to the start of the interview, the participant will be informed they can choose to end the interview at any time, as well as choose not to answer any questions.

F. Will the subjects be presented with materials which might be considered to be offensive, threatening, or degrading? If so, please describe.

The participants will not be presented with materials which might be considered to be offensive, threatening, or degrading.

G. Approximately how much time will be demanded of each subject?

The initial recruitment survey will take less than 5 minutes to complete. The interview process will take approximately 60 minutes.

H. Who will be the subjects in this study? How will they be solicited or contacted? Provide an outline or script of the information which will be provided to subjects prior to their volunteering to participate. Include a copy of any written solicitation as well as an outline of any oral solicitation.

The participants will be instructional designers who have experience with the phenomenon under investigation. Participants will be solicited via a social media post that includes a link to the recruitment survey (see attachment for the social media post and recruitment survey). The social media platform will be LinkedIn because it is a professional networking site where instructional designers participate in communities of practice.

I. What steps will be taken to insure that each subject's participation is voluntary? What if any inducements will be offered to the subjects for their participation?

The initial recruitment survey and the informed consent forms will explain that participation in the study is voluntary. No compensation will be provided to partcipants.

J. How will you insure that the subjects give their consent prior to participating? Will a written consent form be used? If so, include the form. If not, explain why not.

An informed consent form will be included at the beginning of the recruitment survey. The participant must provide consent to continue with the survey. Individuals who are selected to interview will be asked to review and sign a separate informed consent form prior to the interview. See attachments for the informed consent forms.

K. Will any aspect of the data be made a part of any permanent record that can be identified with the subject? If so, please explain the necessity.

No, no data will be made a part of any permanent record that can be identified with the participants. If an individual is selected to interview, their name will be coded to protect their anonymity (e.g., Jane Doe will be changed to Participant A). The coding matrix will be permanently deleted once the study concludes.

L. Will the fact that a subject did or did not participate in a specific experiment or study be made part of any permanent record available to a supervisor, teacher, or employer? If so, explain.

No, no permanent records will be kept for any subjects who choose not to participate in the study or who withdraw from the study.

M. What steps will be taken to insure the confidentiality of the data? Where will it be stored? How long will it be stored? What will be done with the data after the study is completed?

The recruitment survey responses will be stored in SurveyMonkey, which is password-protected. All interview data will be reported in aggregate; the results of the study will not contain information that will personally identify participants. The interview recordings and the participant coding matrix will be stored in a password-protected location and will be permanently deleted when the study concludes.

N. If there are any risks involved in the study, are there any offsetting benefits that might accrue to either the subjects or society?

There are no known risks involved in the study that offset any benefits.

O. Will any data from files or archival data be used? If so, please describe.

No data from files or archical data will be used.

Appendix E. Baker University IRB Approval

BAKER UNIVERSITY
Own Confidence

Baker University Institutional Review Board

March 23rd, 2021

Dear Stephen Naso and Wendy Gentry,

The Baker University IRB has reviewed your project application and approved this project under Expedited Status Review. As described, the project complies with all the requirements and policies established by the University for protection of human subjects in research. Unless renewed, approval lapses one year after approval date.

Please be aware of the following:

1. Any significant change in the research protocol as described should be reviewed by this Committee prior to altering the project.
2. Notify the IRB about any new investigators not named in original application.
3. When signed consent documents are required, the primary investigator must retain the signed consent documents of the research activity.
4. If this is a funded project, keep a copy of this approval letter with your proposal/grant file.
5. If the results of the research are used to prepare papers for publication or oral presentation at professional conferences, manuscripts or abstracts are requested for IRB as part of the project record.
6. If this project is not completed within a year, you must renew IRB approval.

If you have any questions, please contact me at npoell@bakeru.edu or 785.594.4582.

Sincerely,

Nathan Poell, MLS
Chair, Baker University IRB

Baker University IRB Committee
Sara Crump, PhD
Nick Harris, MS
Christa Manton, PhD
Susan Rogers, PhD

Appendix F. Interview Invitation Email

Subject Line: Research Study: Interview Invitation

Hello [Name],

You are receiving this email because you recently completed a survey regarding training evaluation. In that survey, you indicated your interest in participating in an interview. I would be pleased to speak with you and am attaching information about the study and interview process. This attachment also includes an informed consent form for your review and signature.

I'd love to schedule your interview in the next day or two. Can you please reply back with the following?

1. A copy of your signed consent form.
2. A preferred date and time for the interview. Here are a few options (I'm happy to provide more if necessary):
    a. Date/Time Option 1
    b. Date/Time Option 2
    c. Date/Time Option 3
    d. Date/Time Option 4

    Note: we will be using Zoom for the interview, and while I will have my camera enabled so you can see me, you can choose to enable or disable your camera – whatever you are most comfortable with is fine with me.

Please feel free to reach out with any questions. Thank you very much for your participation!

Best,

Stephen Naso
Ed.D. Candidate
Baker University

[Attachment: Interview Process and Informed Consent Form]

Interview Process and Informed Consent Form

1. Background

One common training program evaluation technique involves distributing questionnaires at the end of a training program to measure the trainees' reactions to the program. These questionnaires typically include a mix of closed-ended and open-ended questions and are commonly referred to as Level 1 evaluations, smile sheets, happy sheets, reaction sheets, or reactionnaires. Examples of these questions are as follows:

Closed-ended Questions:	**Open-ended Questions:**
• Did the program meet the stated objectives? • Would you recommend the program to your peers?	• What about the program did you find most valuable? • How could the program be improved?

2. Interview Process

The researcher will schedule the interview via Zoom at a time that is convenient for you. Additionally, the researcher will ask you to review and return a signed copy of the informed consent form (below) prior to the interview occurring. Once you are ready to begin, the researcher will ask a series of questions about the research topic. The interview will take no more than 60 minutes.

3. Informed Consent Form

Purpose
The purpose of this research study is to help uncover how learning and development practitioners analyze the written comments received on training feedback forms.

Participation
You are invited to participate in the interview portion of this study because you completed a survey that indicated your interest in an interview. Additionally, you met the criteria established by the researcher. Your participation in this research study is completely voluntary; you may choose not to participate. If you choose to participate, you may withdraw at any time. If you choose not to participate or if you withdraw from the study at any time, you will not be penalized.

Benefits, Risks, and Discomforts
This study adds to our understanding of how learning and development practitioners analyze qualitative reactions. There are no known risks or discomforts involved with your participation in this study. There are no direct benefits to you as a participant in this study.

Compensation
Once you complete the interview in full, you will receive a $25 Amazon gift card to compensate you for your participation.

Confidentiality
The results of this study will be used for scholarly purposes only. To help protect your confidentiality and anonymity:

- A secure platform (i.e., Zoom) will be used to record and transcribe your interview.
  - Your last name will not be used while the interview is being recorded.
  - A transcript of the interview will be stored in a password-protected electronic format that is only accessible to the researcher, his major advisor, and his research analyst.
  - Once the study concludes, the recording will be permanently deleted.
- All data will be reported in aggregate; the results of the study will not contain information that will personally identify you.
  - Your full name will be coded to maintain anonymity (i.e., Jane Doe will be coded as Participant A, John Doe will be coded as Participant B, etc.). The coding matrix will be stored in a password-protected electronic format that is only accessible to the researcher.
  - Once the study concludes, the coding matrix will be permanently deleted.

Questions About the Study
If you have any questions about this research study, you may contact the researcher or the researcher's major advisor.

Researcher:
Stephen Naso, Ed.D. Candidate
School of Education, Baker University
stephenjnaso@stu.bakeru.edu

Major Advisor:
Wendy Gentry, Ph.D.
School of Education, Baker University
wendy.gentry@bakeru.edu

Note: this study has been approved by Baker University's Institutional Research Board (IRB) for research involving human subjects.

Consent
By signing below, you indicate that you have read this informed consent form, that you understand it, and that you choose to participate in the interview portion of this study.

_____ _____
Signature **Date**

_____
Printed Name