

Summer 8-2022

An Instrumental Case Study of Confirmative Evaluation in the Continuous Improvement of Online Remote Leaders' Work

Chad Lawrence McLane
Old Dominion University, chad.mclane@gmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/stemps_etds



Part of the [Instructional Media Design Commons](#), [Performance Management Commons](#), and the [Systems Science Commons](#)

Recommended Citation

McLane, Chad L.. "An Instrumental Case Study of Confirmative Evaluation in the Continuous Improvement of Online Remote Leaders' Work" (2022). Doctor of Philosophy (PhD), Dissertation, STEM Education & Professional Studies, Old Dominion University, DOI: 10.25777/zcxn-gf46
https://digitalcommons.odu.edu/stemps_etds/130

This Dissertation is brought to you for free and open access by the STEM Education & Professional Studies at ODU Digital Commons. It has been accepted for inclusion in STEMPS Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

**AN INSTRUMENTAL CASE STUDY OF CONFIRMATIVE EVALUATION IN THE
CONTINUOUS IMPROVEMENT OF ONLINE REMOTE LEADERS' WORK**

by

Chad Lawrence McLane
B.A., December 2004, Brigham Young University-Idaho
M.A., May 2007, Marquette University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

INSTRUCTIONAL DESIGN AND TECHNOLOGY

OLD DOMINION UNIVERSITY

August 2022

Approved by:

Dr. John Baaki (Director)

Dr. Noah Glaser (Member)

Dr. Joel Drake (Member)

ABSTRACT

AN INSTRUMENTAL CASE STUDY OF CONFIRMATIVE EVALUATION IN THE CONTINUOUS IMPROVEMENT OF ONLINE REMOTE LEADERS' WORK

Chad Lawrence McLane
Old Dominion University
Director: Dr. John Baaki

Confirmative evaluation is often noted as an important element of models of Human Performance Technology and Performance Improvement, but there exists confusion around what is and what is not confirmative evaluation. A significant issue limiting the use of confirmative evaluation is the Performance Improvement field's inability to clearly classify confirmative evaluation in terms of its purpose and scope. Additionally, the performance improvement field lacks sufficient examples of confirmative evaluation in the literature necessary to adequately define confirmative evaluation and demonstrate its use. Without sufficient examples of confirmative evaluation, practitioners of performance improvement are left without a clear path of how to conduct confirmative evaluation in ways that are contextually relevant and cost effective. Additionally, those who train instructional designers and performance improvement practitioners lack the tools necessary to help students of the field learn about and use this vital aspect of evaluation.

Through the conceptual framework of systems theory and change management, this instrumental case study reviewed the literature around confirmative evaluation to identify the confusion around confirmative evaluation and then presented a case of confirmative evaluation applied in a context of change management and continuous improvement in online higher education. This case used job descriptions of remote leaders' work to determine the degree to which an organizational restructuring had been implemented as it had been intended and had

influenced remote leaders' actual work. Time study methodology and matrix sampling were employed to observe remote leaders' work. The data from the time study was compared against remote leaders' job descriptions and institutional surveys gauging satisfaction among those served by these remote leaders. The difference between the expected work and the observed work demonstrated the degree to which remote leaders had embraced the new duties given to them in the organizational restructuring and identified key gaps in their collective performance. Together, this instrumental case study demonstrated that confirmative evaluation plays a linking or bridging role in continuous improvement processes by moving the process from summative evaluation to a new phase of formative evaluation.

The study then examined the results of the case to consider how it clarified the field's understanding of confirmative evaluation. Key findings included the recognition that confirmative evaluation is distinct from other forms of evaluation because of its purpose and that purpose and timing are intertwined and inform one another in evaluation efforts. It also identified where confirmative evaluation fits within and can improve performance improvement and change management models. The study ended with a discussion of challenges faced when conducting confirmative evaluation and implications for future studies of confirmative evaluation.

Copyright, 2022, by Chad Lawrence McLane, All Rights Reserved.

This dissertation is dedicated to my family, especially Mindy. I can't express here what your support most of all has meant to me. This would not have happened without you.

ACKNOWLEDGMENTS

There are many people who have contributed to the successful completion of this dissertation. First, I would like to thank Dr. John Baaki, Dr. Noah Glaser, and Dr. Joel Drake. Their guidance and insight helped focus this work immensely. I would also like to thank Steve Stokes for their assistance in data collection. Thanks goes to Dr. Pamela Fly for encouraging me to pursue a PhD when I had every reason not to. I also thank the BYU-Idaho Online Learning family generally for their support of continued personal growth and interest in improving the lives of those they work with and those they serve throughout the world.

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES	x
INTRODUCTION	1
Context of the Case	2
Conceptual Framework	3
Current Study	8
Organization of the Study	13
LITERATURE REVIEW	15
Systems Theory	15
Change Management and ISPI's HPT model	17
Job Descriptions and Performance in Systems	19
Confirmative Evaluation	20
Conclusion	39
METHODOLOGY	41
Instrumental Case Study Design	41
Identification of the Case	42
Sources of Evidence	48
Instruments	52
Time Study Data Collection Procedures	54
Limitations	57
Conclusion	59
RESEARCH FINDINGS	60
Research Question 1	61
Research Question 2	90
Research Question 3	97
Conclusion of Findings	101
Limitations & Future Research	102
Discussion & Implications	106
Is confirmative evaluation different from formative or summative evaluation?	108
What role does confirmative evaluation play in performance improvement and continuous improvement models?	114
What challenges are faced when conducting confirmative evaluation?	124
Conclusion & implications for future study	128

REFERENCES 131
VITA 142

LIST OF TABLES

Table	Page
Table 1 Evaluation Conceptualization Framework.....	35
Table 2 Online Instructor and Course Group Leader Demographics.....	46
Table 3 Course Group Size	47
Table 4 Time Study Response Rate	62
Table 5 Comparison of Remote Leader Role & Size Weekly Average.....	64
Table 6 Comparison of Time Spent by Role, Institution, and Group Size	65
Table 7 Comparison of Contracted Time with Average Time for OCRs	66
Table 8 Comparison of Contracted Time with Average Time for ACLs.....	68
Table 9	71
Table 10 Micro Group OCR Reported Time Week by Week.....	72
Table 11 Small Group OCR Reported Time Week by Week	75
Table 12 Micro Group ACL Reported Time Week by Week	78
Table 13 Small Group ACL Reported Time Week by Week	80
Table 14 Fall 2021 Course Group Survey Results.....	84
Table 14 continued.....	85
Table 15 Comparison of Percent of Course Lead Observations of Remote Leader Activity	86
Table 16 Comparison of Remote Leader Time by Role & Size of Group.....	87
Table 17 Instructor Satisfaction with Course Groups	92
Table 18 Comparison of Course Lead Perception of ACL Value Winter 2021 and Fall 2021 ...	93
Table 19 Gaps in Remote Leader Performance and Areas of Investigation	99
Table 20 Group N and Average Weekly Response Rates.....	104
Table 21 Evaluation Conceptualization Framework Applied	118

LIST OF FIGURES

Figure	Page
Figure 1 ISPI HPT Model (Dessinger et al., 2012).....	18
Figure 2 Marker et al.'s (2014) Spiral HPI Framework, with Confirmative Evaluation.....	28
Figure 3 Week by Week Comparison of Remote Leader	72
Figure 4 OCR Micro Group Principal Duties Week by Week	75
Figure 5 OCR Small Group Principal Duties Week by Week.....	77
Figure 6 ACL Micro Group Principal Duties Week by Week.....	80
Figure 7 ACL Small Group Principal Duties Week by Week	82
Figure 8 Moseley & Hastings Intervention Implementation Process Model, Revised.....	118

CHAPTER I

INTRODUCTION

Change management is a process of identifying needs, implementing interventions, and actively influencing the adoption of change and innovation to achieve the desired results (Van Tiem et al., 2012). In complex organizations with over a decade of institutional history and many partners who hold significant stake in the organization's activities, change management is challenging. Structural, political, human resource, and symbolic anchors must be adapted as old ways are replaced with new ways that move the organization into the future (Bolman & Deal, 2017). Changes cannot successfully occur in a vacuum as they are all parts of a larger system full of people, processes, and subsystems with their own needs that must be balanced (Lewin, 1946; von Bertalanffy, 1972). Change must be intentionally managed and often requires prolonged attention. But how do we know if implemented changes have had the desired, lasting effects?

Confirmative evaluation, while often overlooked, confused, or ignored in the literature and the work of instructional designers, is a valuable measure of successful change and innovation (DeVaughn & Stefaniak, 2020a). The importance of formative and summative evaluation is well established among designers and educators generally. Less familiar but equally valuable is confirmative evaluation. Where formative evaluation improves products and processes while they are being enacted and summative evaluation identifies to what degree the developed products and processes function as they are intended, confirmative evaluation considers if products and processes continue to produce the desired effects over time and evaluates the overall effect of an intervention (Dessinger & Moseley, 2015). Confirmative evaluation can be a costly endeavor, but one that solidifies the value of enacted changes (Dessinger & Moseley, 2015).

A significant issue limiting the use of confirmative evaluation is the Performance Improvement field's inability to clearly classify confirmative evaluation in terms of its purpose and scope. While it is generally accepted that confirmative evaluation is a valuable element of performance improvement, the performance improvement field lacks the literature and examples necessary to truly understand and utilize confirmative evaluation. Part of the challenge the field has had in applying confirmative evaluation has been that many efforts for defining evaluation focus on outputs, timing, or tools used for the evaluation. When evaluation is instead first defined by its purpose, a different typology emerges that locates confirmative evaluation alongside formative and summative evaluation within a continuous improvement environment.

This study presents an instrumental case study of confirmative evaluation applied in a context of change management in online higher education. Its purpose is to help fill the current gap in literature around confirmative evaluation and to examine how confirmative evaluation functions in a continuous improvement model. It also addresses common concerns of cost and feasibility of confirmative evaluation by using industrial time study methodology to quantify the work and value of remote leaders.

Context of the Case

The online learning department showcased in this study is in a private university located in the intermountain west. This particular online learning department employs over 2,000 online instructors to serve over 60,000 students enrolled in online courses at three institutions in a shared education system. This department had recently undergone an organizational restructuring that simplified instructor teams and reporting lines. This restructuring started in the Winter 2021 semester and was fully enacted in the Spring 2021 semester. While this change brought many benefits of simplicity and cost savings, it worked against a decade of organizational philosophy,

policies and procedures, and organization-specific software development. Because this restructuring was contrary to so much organizational history and context, the migration to the new organizational model required substantial planning, messaging, training, and change management to ensure all stakeholders understood the changes, how those changes affected them, and their new roles and responsibilities. With the start of the Fall 2021 semester, all online courses were in the new organizational model for over four months and the university's Online Learning department was at a stage in this metamorphosis where it needed to evaluate if the changes enacted thus far have resulted in the desired changes in remote leader activity.

To this end, this case used confirmative evaluation to determine how well the organization had abandoned the old model of instructor organization and support in favor of the new model by examining the activity of remote leaders in the new organizational model. In the old model, remote leaders' primary function was to provide support and mentoring for other online adjunct instructors. Under the new organizational model, remote leaders are meant to focus their efforts on course design and delivery improvement. If remote leaders had continued to function as they had before the organizational restructuring, Online Learning had done little more than put old wine in new bottles and failed in its efforts to change the focus of remote leaders' primary work and subsequently the nature of the organization at large.

Conceptual Framework

It is in this context of deep organizational change that this instrumental case study was conducted. Instrumental case studies are useful for providing insight in an issue or to refine theory (Baskarada, 2014) and differ from intrinsic case studies in that intrinsic case studies examine a case for the uniqueness of that case while instrumental case studies use a case to understand a larger phenomenon by observing it within a case (Harling, 2002; Mills et al., 2010).

Instrumental case studies are conducted to provide insight into a larger issue or redraw a generalization (Stake, 2003). In this instance, the instrumental case study was used to examine how confirmative evaluation can be used in a real continuous improvement context, unlike previous case studies of confirmative evaluation placed in hypothetical situations (Dessinger & Moseley, 2015). This case study was founded on a conceptual framework of systems thinking, change management, and confirmative evaluation. A brief review of these concepts will situate this study in the broader context of each of these areas and highlight how it informed this work while also providing greater research clarity on these subjects.

Systems Theory

Systems theory was the first conceptual foundation for much of the work in this study. Systems theory recognizes connections of individuals within larger structures (Boulding, 1956). To understand an entire structure, we must understand its parts and processes *and* their interconnections and relationships. Only by linking the disparate elements of a system do we begin to grasp the entire system and its meaning (von Bertalanffy, 1972). Systems consist of subsystems, processes, outputs, inputs, suprasystems, and the environment (Richey et al., 2011), and every system is a constant balance of each of these elements with one another. Instructional designers work within the context of these elements, and their work depends on their ability to understand and adapt to these layers.

Change Management

This project was also grounded in the study of change management. Change management is an important task for instructional designers and performance improvement professionals (Brigance, 2011). Change management is a process of ensuring an intervention has the best chance at being adopted or influencing individuals or organizations and is best considered in a

systems conceptual framework. In this light, changes are the result of system disruption that cause disequilibrium and complex problems for the system (Tamim, 2020). Understanding how a system operates places us in a better position to enact strategies that will change the institution (Heinich, 1985; Tamim, 2020), and systems approaches are necessary for truly disruptive change (Ellsworth, 1997; Fullan, 1996; Tamim, 2020; Tessmer & Richey, 1997). Christensen and Eyring (2011a) identified that online learning presented serious competition and disruption in the higher education system. To help address the disruption presented in change, a change manager must align innovations with the institution's needs and resources to ensure successful change (Langdon, 2000; Rogers, 2003).

According to Christensen and Eyring (2011a), traditional higher education must adapt to the changes presented by online learning to avoid serious, even fatal, disruption and to leverage the benefits of online learning that can expand the traditional university's capacity. Online education is challenging for educational leaders and requires competence in handling change (Christensen & Eyring, 2011b; Tamim, 2020). The disruptive innovation online learning presents higher education must be addressed in higher education's political, economic, and academic power centers seeking to maintain the status quo (Beaudoin, 2016). Bolman and Deal (2017) additionally argued that strategies for improving organizations lie in how changes to the organization are framed in structural, human resource, political, and symbolic lenses. Leadership in online education is more about managing change through these lenses than it is managing technology and requires agility, interprofessional leadership, civility, inclusiveness, and strategic communication (Menon & Suresh, 2020; Thompson & Miller, 2018).

Effective change management will ensure innovations have long-lasting impact on the system in which they are enacted. This requires the thorough study of an implementation to

identify desired changes and then determine if a change had an intended effect (Fullan, 1996). While many models of change management exist (Waterman et al., 1980; Kotter, 1995; Imai, 1997; Antony et al., 2017), this study was grounded in the International Society of Performance Improvement's Human Performance Technology Model (Van Tiem et al., 2012) for its clear emphasis on confirmative evaluation.

Confirmative Evaluation

Confirmative evaluation can help with this important step of determining if a change had its intended effect and represents the third conceptual foundation for this study. Confirmative evaluation is a vital but often overlooked component of evaluation that can be used along with formative and summative evaluation (Giberson et al., 2006). Formative evaluation informs design decisions while in production stages, and summative evaluation ensures that a product initially performs as it was intended. Formative and summative evaluation focus on immediate results of a change, but they fail to support long-term programs or systemic approaches that are necessary for adoption of truly disruptive innovations. Confirmative evaluation fills this gap by examining the long-term impact of a product or change beyond initial launch of the intervention. It attempts to consider the total impact of a change or innovation. Where summative evaluations look backwards, confirmative evaluation looks forward and attempts to help planners make decisions on the future (Dessinger & Moseley, 2015) while building from where formative and summative evaluation leave off and can be used to judge the continuing worth or merit of a program (Dessinger & Moseley, 2015; Moseley & Solomon, 1997).

Confirmative evaluation's value to change management can be immense. Confirmative evaluation determines the degree to which an innovation has enacted the desired change over an extended period and can reveal how well a change in a system has settled that system into a new

equilibrium. Confirmative evaluation is best undertaken three to fourteen months after a change has been enacted (Dessinger & Moseley, 2015), extends traditional evaluation measures to measure long term performance, and can assist organizational change processes by identifying further needed changes and the degree to which changes have been institutionalized by an organization (Giberson, et al., 2006). Confirmative evaluation is inherently interested in a training or intervention's value and tries to identify, explain, confirm, or justify the continued value of that training or intervention over time (Van Tiem et al., 2000), often by using cost analysis methods (Dixon, 1990; Guerra-Lopez, 2008; Jackson, 1989; Phillips et al., 2006).

Unfortunately, confirmative evaluation is grossly underutilized. One reason for this is the investment gap between adopting a change and implementing it (Fullan, 1996). According to Fullan (1996) identifying the need for a change and how to achieve that change is a relatively easy sell to organizational leaders. It is much harder to convince them of the value of evaluating long enacted changes to ensure the changes have had the desired effects. This is especially difficult when the cost of confirmative evaluation for decisions that have already been made is relatively high and do not always clearly lead to increased performance or bottom-line improvement (Guerra-Lopez, 2008; Jackson, 1989; Williams et al., 2011).

Because of the relative difficulty in conducting confirmative evaluation, subsequent study of this vital evaluation component is lacking (Dessinger & Moseley, 2015; DeVaughn & Stefaniak, 2020a; Giberson et al., 2006). Dessinger and Moseley's (2015) landmark work on confirmative evaluation relied on example cases that, while they demonstrate the value of confirmative evaluation, lack true power because they are situated in hypothetical scenarios like the Oz City Zoo. This dearth of real cases of confirmative evaluation has lasted almost twenty years and has negatively impacted the performance improvement field's ability to embrace

confirmative evaluation. Most recently, DeVaughn & Stefaniak (2020a) found using semi-structured interviews of instructional designers in national defense, higher education, government, and private industry that confirmative evaluation was reportedly used inordinately less than formative and summative evaluation. In their study, 66% of designers reported not conducting confirmative evaluation. DeVaughn and Stefaniak (2020a) further found that confirmative evaluation is challenged by a lack of rich data and institutional apathy towards the importance of confirmative evaluation.

Because of its relative underutilization, there is a considerable gap in the literature regarding application of confirmative evaluation. That gap in the literature further compounds the challenge of understanding and then using confirmative evaluation because the field lacks examples and models of confirmative evaluation. Confirmative evaluation can be a useful tool to support organizational goals and aligns performance improvement interventions with valued outcomes (Marker et al., 2014). The instructional design field's general underutilization of confirmative evaluation hinders its ability to identify the true cost and value of instructional and non-instructional interventions and to justify its benefit to organizations. Revising our understanding confirmative evaluation will help our field better apply concepts of confirmative evaluation and illustrate the value and role instructional designers and performance improvement professionals play in organizations.

Current Study

As noted before, the change in how online instructors were organized and managed was a significant organizational shift that upended a decade of organizational policy and the institutional culture that developed around it. In this instrumental case study, I conducted an example of confirmative evaluation to determine the degree to which the change to a new

organizational model has been enacted. At the start of the Fall 2021 semester, the new organizational model had been in place for four months and was at the ideal time for a confirmative evaluation to determine if the enacted organizational changes have had the desired effects.

Central to the success of the new organizational model is the work remote leaders do in their contracted responsibilities. This model is led by two types of remote leaders: Online Course Representatives (OCR) and Assistant Course Leads (ACL). Their work and the time allotted to that work is substantially different from the old remote leaders' duties and time, and the Online Learning department needed to be certain both types of remote leaders were performing their jobs as they were designed. While the OCR and ACL roles were extremely similar, they were different in that ACLs have authority OCRs do not to make fixes in the master course. This difference, along with the different sizes of course groups ACLs and OCRs each serve, makes each remote leader role unique. Understanding the similarities and differences of these roles as they are currently deployed would help the organization determine where each role can be best deployed within the organization.

Additionally, Online Learning needed to ensure that the designed job and contract load adequately fit the ACL and OCR roles. If the job as it was designed requires too much time of OCRs and ACLs, they would likely suffer from burnout and exhibit poor performance. Likewise, if ACLs and OCRs went beyond what is expected of them, they could create unnecessary dependencies among the instructors they serve and distract ACLs and OCRs from their other vital job duties. Job descriptions for both roles were an important element to determining if the desired changes have in fact been effectively implemented. From this information, the

organization's next steps in cementing the new organizational model and needed revisions to the remote leader job descriptions and trainings would become clear.

In this context, this study sought to answer the following questions:

1. How can confirmative evaluation be used to determine the degree to which organizational changes have been institutionalized?
2. To what degree can confirmative evaluation determine the effects of changes in continuous improvement efforts?
3. How does data from confirmative evaluation support continuous improvement efforts to reinforce organizational changes?

Methodology

The research questions above were answered using a time study methodology and a matrix sampling procedure to collect the necessary data. Time study methodology is common in manufacturing and industry to evaluate worker performance and establish baseline performance expectations. It functions by dividing processes or work into its sub-processes and actors, evaluating the behavior of the sub-processes and actors, and summing all parts of the process to obtain an overall system timing (Yazdi et al., 2019). The time study in this case examined how ACLs and OCRs spent their time over the 15 weeks of a semester at this institution compared with their job descriptions to identify gaps in actual versus expected performance.

Time studies over such a long duration of time could face considerable challenges like ensuring an adequate number of respondents complete the entire survey used to collect the needed data over the entire semester, or on the opposite end, producing too much data to compute efficiently (Gonzalez & Eltinge, 2007; Thomas et al., 2006), so a matrix sampling methodology (Thorndike & Thorndike-Christ, 2010) was employed to subdivide the survey into

smaller data collection chunks to track the activity of the 351 ACLs and OCRs throughout the Fall 2021 semester. In this project, matrix sampling created a composite view of the remote leadership roles while minimizing the time investment and potential perceived risk for any single member of the study and reducing administrative cost (Childs & Jaciw, 2002; Kaplan & Su, 2016). This study was unique because while most matrix sampling efforts split long surveys among many respondents (Gonzales & Eltinge, 2007; Thomas et al, 2006), this study used the same survey that would be applied many times over many months among many respondents. Under this conceptualization, this time study was a survey of the different activities required of course group leaders over each week of the semester, and the weeks of the semester were the grouping logic that organized the sampling.

In this study, remote leaders were organized by course group size and type of contract and divided into 15 survey groups. Each remote leader was surveyed four times in the semester to ensure adequate sampling while not substantially increasing their workload. Results from these surveys were analyzed using univariate statistics (Gonzales & Eltinge, 2007) to identify averages and totals for each week and the weekly averages for each activity, and the results from this time study were triangulated against satisfaction and observation surveys already in use by the organization.

Significance

For the organization, this time study and triangulation with existing survey data demonstrated the degree to which the change to the new organizational model had been adopted. This time study provided the university's Online Learning department with a clear view of how ACLs and OCRs spend their time throughout a typical semester. By developing composite views

of the ACL and OCR roles each week, this study identified how the roles are similar or different over the total course of the semester. With this view, Online Learning could confirm the degree to which the organization had successfully designed and implemented the new organizational model. This information could inform future job design and training activities for new ACLs and OCRs. These data also revealed gaps in the transition to the new leadership model and suggested where the organization must continue to manage the change to the new organizational model.

For the field of performance improvement, this instrumental case study shed light on confirmative evaluation's real power by providing a much-needed real-life example of confirmative evaluation in action in higher education and demonstrated how confirmative evaluation could be conducted in a way that fits organizational need by employing contextually unique data collection methods. Further, it accurately located confirmative evaluation alongside formative and summative evaluation by demonstrating its transitional role in continuous improvement efforts and highlighted the uniqueness of confirmative evaluation compared with formative and summative evaluation. Thereby, it showed that confirmative evaluation is a distinct form of evaluation, and as such deserves more attention than it has previously gained.

Assumptions and Limitations

In conducting this study, several assumptions are made. First, the participants in this study would provide honest and accurate representations of the time they spend in their work as remote leaders. Second, the instruments used would elicit reliable representations of remote leaders' job duties. Third, the amount of time remote leaders spent on different duties change over the course of the semester; if this were not the case, a week-by-week study would not be necessary.

Additionally, this study came with some limitations. First, because this instrumental case study centered on the use of a time study in an organization, there was potential in this study that participants would not accurately represent their work if they feared their responses may reflect poorly on themselves. Transparently protecting participants' anonymity helped address this limitation. Tied to this, this study relied on self-reported data that could be corrupted through inconsistent reporting by participants, and care was taken to ensure participants report honestly and accurately. Additionally, course groups with more than 20 instructors were not represented in this study because so few of these larger course groups exist that it made protecting these participants' anonymity impossible. This limitation impacted the inferences the organization could make about its largest course groups but given there were far fewer of these large course groups than the other size groups, this limitation's impact was minimal. This study was further limited by only exploring a single semester of ACL and OCR work. Additional semester time studies would increase the reliability of the data and its applicability throughout the year but lied outside the scope of this project. Finally, this study did not distinguish among academic department or college. A comparison across academic disciplines may have revealed leadership differences but was outside the scope of this study.

Organization of the Study

This study was organized into five chapters. Chapter One introduces the study by explaining the issues around confirmative evaluation, briefly describing the context of the case, and establishes the rationale, conceptual framework, and general approach of the study. Chapter Two reviews the relevant literature that guides this work and locates it in the realms of systems theory, change management, and evaluation theory. Chapter Three outlines the research methodologies employed in this study and introduces and explains the participants, instruments,

data collection procedures, and analysis procedures. Chapter Four presents the results of the research. Chapter Five is a discussion of the results of the study, their implications, and conclusions drawn from these results and implications. With the introduction of the study context complete, it is time to move to the review of relevant literature.

CHAPTER II

LITERATURE REVIEW

Understanding confirmative evaluation's value requires it be placed within its larger context of evaluation generally and evaluation's role in change management. Doing so helps us understand why it is mentioned broadly in the literature but seldom undertaken. As a vital component of performance improvement, it can connect change initiatives to an organization's valued outcomes. But for us to really understand confirmative evaluation, we must begin by seeing it in the contexts of systems theory and change management. This chapter will begin with an explanation of essential elements of systems theory and introduction to the International Society of Performance Improvement's Human Performance Technology Model (Van Tiem et al., 2012) of change management. It next explores job descriptions as they are the foundation for the confirmative evaluation conducted in this instrumental case study. This chapter will then explore the concept of confirmative evaluation and where it fits in change management and in relation to other types of evaluation. It concludes with an argument for what is lacking in the current literature about confirmative evaluation and a call for expanding our conceptualization of confirmative evaluation's methods and the relevance for performance improvement practitioners and the organizations they serve.

Systems Theory

Systems theory recognizes connections of individuals within larger structures (Boulding, 1956). To understand an entire structure, we must understand its parts and processes *and* their interconnections and relationships. Only by linking the disparate elements of a system do we begin to grasp the entire system and its meaning (von Bertalanffy, 1972). Systems consist of subsystems, processes, outputs, inputs, suprasystems, and the environment (Richey et al., 2011).

Von Bertalanffy (1972) explained that general systems theory is a study of the interrelations of parts of a system and that form, order, and hierarchy are essential to understanding a system. Every system is constantly balancing each of these elements with each other and are self-stabilizing (Laszlo, 1996) through positive and negative feedback as they seek dynamic equilibrium, or homeostasis (Skyttner, 2001). When homeostasis is achieved, systems are balanced and become frozen in their state and require significant effort to unfreeze, change, and refreeze (Lewin, 1946). Any change introduced into a system requires adjustments throughout the entire system. Fullan (1996) identified the systemic nature of critical factors of change in the characteristics of the change, local characteristics that influence the efficacy of a change, and external factors that influence the change. Failure to enact a change systemically increases the likelihood that the system will revert to its original state. Only through systemic change, not through a single change, can systems be permanently shifted (Ellsworth, 1997).

A challenge in effecting change in systems is clearly defining the system and its components. Von Bertalanffy posited that parts of systems overlap and that the spatial boundaries around a thing are blurry. What makes a system definable is what is cohesive, or the interactions among component elements of a system. Adams et al. (2020) argued that a challenge in taking a systems approach is that the boundaries of an open system are dynamic rather than spatial. Because they are dynamic, they are indistinct, making it difficult to identify what belongs to a system and what does not. Only by identifying hierarchies or boundaries do we narrow the system sufficiently to focus adequately on the assessment at hand to understand the thing we have bounded. Adams et al. (2020) agree with Giberson (2015) that it is impossible to analyze all elements of a system, and that boundaries are needed to develop deep knowledge. Because of the dynamic nature of systems and the blurriness of boundaries, general systems theory allows for

multiplicity of approaches (equifinality) and is inherently messy. For a practitioner, this means it is acceptable to take many different approaches to find solutions.

Change Management and ISPI's HPT model

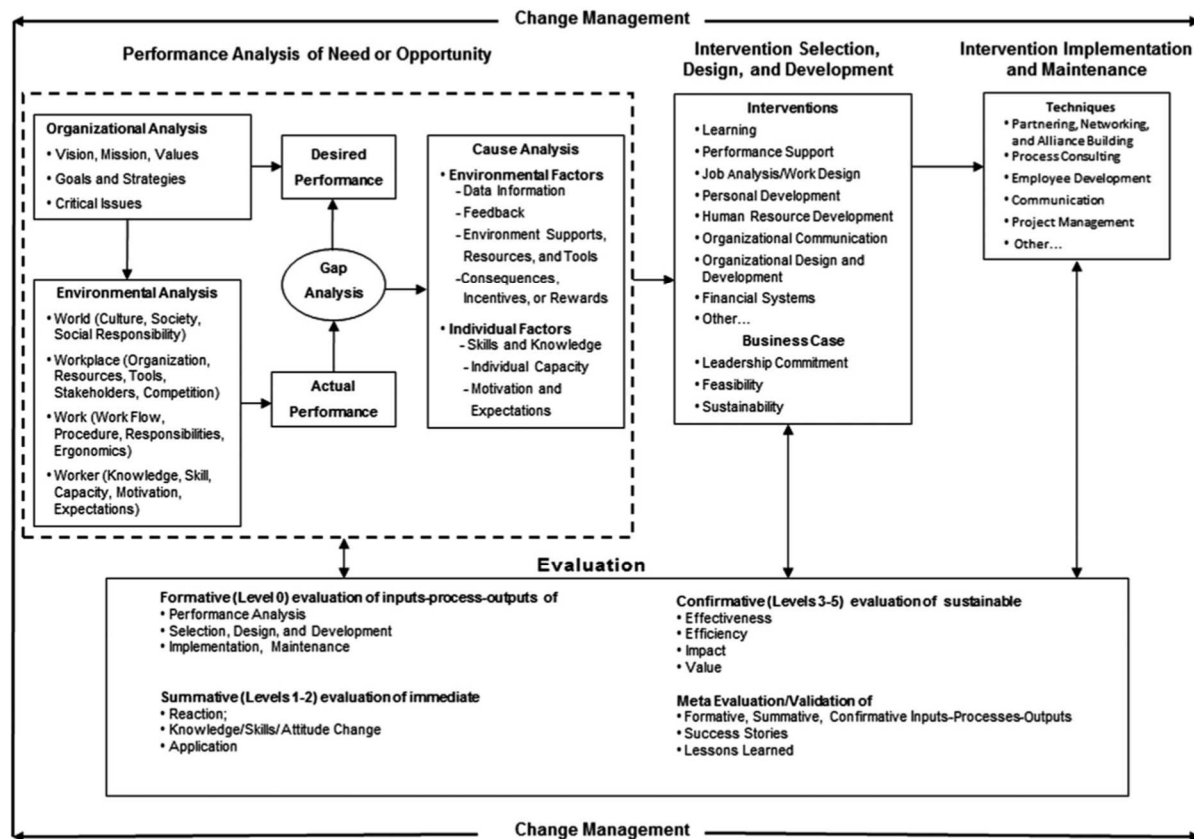
Understanding how a system operates places us in a better position to enact strategies that will change the institution (Heinich, 1985; Tamim, 2020), and systems approaches are necessary for truly disruptive change (Ellsworth, 1997; Fullan, 1996; Tamim, 2020; Tessmer & Richey, 1997). Regardless of an organization's structure, change management requires thorough study of the implementation to identify desired changes and then determine if a change had an intended effect (Fullan, 1996). Performance improvement seeks solutions that solve the whole problem, which typically requires integrating interventions across the entire organization (Lewin, 1946; Pershing, 2006). To create a comprehensive solution, performance improvement professionals must look at the organization systemically and intentionally manage change to increase the potency of interventions.

While there are many approaches to change management like McKinsey's 7-S Framework (Waterman et al., 1980), Kotter's 8-Step Model (Kotter, 1995), Imai's (1997) Kaizen, and Lean Six-Sigma (Antony et al., 2017), this study intentionally focuses on Dessinger et al.'s (2012) Human Performance Technology model (figure 1). This model replaced the Van Tiem et al.'s (2000) model and represents a significant shift in HPT thinking regarding the importance of change management by placing all performance improvement activity in the context of change management. For Van Tiem et al. (2012), enacting change is the reason for HPT's existence and is a constant requirement for organizations. This model was intended to be both linear and iterative. By being presented as a linear model, it provides a framework from

which new performance improvement professionals can work and explain their work to their clients.

Figure 1

ISPI HPT Model (Dessinger et al., 2012)



Its use of double arrows to link each area to evaluation demonstrates that they inform one another and that practitioners move back and forth throughout the model as context dictates. It also does not distinguish or assign a type of evaluation to a specific stage, noting that all forms of evaluation should be considered throughout the entire process (Hastings, 2009). The model is meant to be systemic and flexible, providing a framework that performance improvement practitioners of all levels of experience can use to consider and explain how they do their work.

Criticism of the ISPI HPT model includes that the linear nature of the process is archaic and inaccurately represents how performance improvement professionals really work (Marker et al., 2014; Czeropski & Pembroke, 2017). It has been interpreted as being less responsive to contextual influences during intervention projects than Agile, SAM, LLAMA, and Lean Six Sigma but still a significantly useful model to be used in tangent with other design models (Alarifi & Alamri, 2014; Czeropski & Pembroke, 2017). Even with these criticisms, it is most applicable to this study because it clearly identifies the role evaluation, particularly confirmative evaluation, plays in change management.

Job Descriptions and Performance in Systems

Equally important as change management models in organizational change are clear performance expectations of individuals and groups. Without a clear understanding of what individuals or groups are expected to do and actually do in their roles, organizational change efforts can easily go amiss by overlooking vital elements of the job or organization (Mader-Clark, 2013). Job descriptions are job aids that define or bound the duties of a particular set of individuals within a system and can be invaluable when evaluating and managing performance (Mader-Clark, 2013; Kshatriya, 2016). Job descriptions document the idealized expectations of a position (Carliner et al., 2015). Mohamed and Hossny (2020) found that job descriptions influence role clarity, which positively correlates with job satisfaction. Additionally, strong job descriptions contribute to better work environments, stable workplace relations, and more coordination across the organization (Pató, 2017). Ashraf (2017) found that the quality of job descriptions influences organizational effectiveness and efficiency by impacting hiring, direction for work to be performed; and organizational KPIs to evaluate workers' performance. In short,

job descriptions can be powerful tools for defining and evaluating the work of specific individuals or groups.

A challenge related to job descriptions is that descriptions may not match the reality of the job (Carliner et al., 2015). Barbouletos (2011) defined job discrepancy as the difference between job descriptions and the actual work performed and found that job discrepancy impacts stress and tension in the workplace. Wakefield et al. (2009) concluded that ill-defined job descriptions create difficulties in defining boundaries of work and establishing accountability and responsibility. Because of the potential value of strong job descriptions and the difficulties associated with ill-defined job descriptions, the quality and accuracy of job descriptions within an organization can carry significant utility when performance improvement professionals are designing and managing initiatives and offers useful means to measure value created by change initiatives.

Confirmative Evaluation

Confirmative evaluation is the process that helps performance improvement practitioners analyze and place value on the total change. Confirmative evaluation was first identified by Misanchuk (1978a, 1978b) as another type of evaluation in addition to Scriven's (1967) formative/summative evaluation dichotomy. Misanchuk argued that the purpose of evaluation is to permit informed decision making and that Scrivner's use of the formative/summative dichotomy mislabels many types of evaluation and causes confusion of what is happening in evaluation. They identify the elements of formative evaluation, including the required durability (repeated application) of a product or process with consistent results. For Misanchuk (1978a) the purpose of formative evaluation is almost always to modify a program under development. Conversely, Misanchuk (1978a) argued that summative evaluation is an evaluation of a finished

product. It asks, "Should this product be implemented?" and its purpose is to demonstrate what a product is capable of. Here, Misanchuk draws clear delimitations between summative and formative evaluation by saying they serve inherently different purposes. According to Misanchuk (1978b), confirmative evaluation is inherently different from both formative and summative evaluation. Unlike formative evaluation, which looks for revision of a product, confirmative evaluation makes no such assumption that a product should even be continued and leaves the option of eliminating the program on the table. In this way, its purpose is not necessarily always to improve a project under development. Conversely, while summative evaluation describes what an object can do, confirmative evaluation fills a gap in purpose the formative/summative dichotomy leaves by examining if the object's current performance justifies its continued support, its redevelopment, or its discontinuance. For Misanchuk (1978c), confirmative evaluation is like summative evaluation in that it makes a "final" judgement on a product's value, but confirmative evaluation is also like formative evaluation in that it looks to revise a product if such action is warranted. They argue that confirmative evaluation is inherently contingent upon the context of evaluation, and that contingent nature makes confirmative evaluation different from both formative and summative evaluation.

Elements of Confirmative Evaluation

Confirmative evaluation examines a training or intervention's value and tries to identify, explain, and confirm or justify the continued value of that training or intervention over time (Van Tiem et al., 2000), often using cost analysis methods (Dixon, 1990; Guerra-Lopez, 2008; Jackson, 1989; Phillips et al., 2006). Formative evaluation informs design decisions while in production stages, and summative evaluation ensures that a product performs as it was intended. Both focuses are on immediate results of a change, but they fail to support long-term programs.

Confirmative evaluation fills this gap by examining the long-term impact of a product or change beyond initial launch of the intervention. Where summative evaluations look backwards, confirmative evaluation looks forward and attempts to help planners make decisions on the future (Dessinger & Moseley, 2015).

One element of context that influences the type of evaluation being performed is the person conducting the evaluation. According to Misanchuk (1978a), formative evaluation is conducted by someone creating a product, implying they have sufficient control over the project to adjust it while it is in development. Misanchuk (1978a) argued that summative evaluation should be performed by someone removed from the product to afford them greater objectiveness. For Misanchuk (1978a), confirmative evaluation blends the two by having someone with the ability to make changes to the product take a highly objective approach to it.

Likewise, Misanchuk (1978b) explained that the timing of evaluation is another important contextual detail that identifies if the evaluation is formative, summative, or confirmative. Formative evaluation is conducted while a product is being created. Summative evaluation occurs after a product is finished but before it is widely adopted and implemented. Confirmative evaluation is conducted after the product has been implemented and used for a period of time. Confirmative evaluation should only be undertaken on a product that has been in place for some time and whose current efficacy is in doubt (Misanchuk, 1978a). While Misanchuk's formative/summative/confirmative trichotomy focuses primarily on the purposes of evaluation, they also simultaneously noted the way the of timing of evaluation influences its purpose (Misanchuk, 1978c). Purpose and timing, for Misanchuk, were interrelated and influenced one another.

After Misanchuk's initial writings in 1978, confirmative evaluation remained essentially unexplored until Hellebrandt and Russell (1993) clarified that the role of confirmative evaluation is to confirm the continued worth of instruction or the competencies of individuals. Hellebrandt and Russell (1993) emphasized the time difference between when formative, summative, and confirmative evaluation occurs, but agreed with Misanchuk that confirmative evaluation seeks to determine if a learner is still competent after some time has passed between learning and evaluation. An undercurrent of this thinking is that confirmative evaluation seeks to identify the change that has occurred in learners based on the implementation of the change or instruction. Hellebrandt and Russell (1993) reinforced Misanchuk's position that another primary difference among the three types of evaluation is the role of the evaluator, and they extended their argument to suggest that a team approach for confirmative evaluation is ideal as it allows for sufficient familiarity with the product to elicit change where necessary and the distance necessary to evaluate conditions as they exist objectively.

Hellebrandt and Russell's (1993) most unique contribution to our understanding of confirmative evaluation was how confirmative evaluation creates a cycle of evaluation with formative and summative evaluation. While they failed to fully explain this concept, they suggested that confirmative evaluation fills a liminal space between summative and formative evaluation that lets evaluators create a continuous improvement cycle rather than a linear evaluation model. Misanchuk (1978a) identified that confirmative evaluation extends evaluation beyond determining if a product needs change during production or if it is meeting its intended effects by asking if it continues to perform as desired some time after implementation is complete. Their explanation places confirmative evaluation in a linear progression chronologically after formative and summative evaluation has been conducted. Hellebrandt and

Russell (1993) noted that confirmative evaluation helps evaluators loop back from summative evaluation back into formative evaluation. This looping role of confirmative evaluation creates a cycle of evaluation that promotes continuous improvement and iterative revision over long periods of time.

After Hellebrandt and Russell's (1993) resurrection of confirmative evaluation, Patton (1996) argued that the formative and summative dichotomy forces us to ignore other roles of evaluation and that the formative-summative dichotomy is insufficient because it focuses on evaluation findings, not the usefulness of the evaluation process. By thinking about the process of doing evaluation, we find many purposes for evaluation beyond formative and summative. It is sometimes impossible to separate the impact of the evaluation process from the program intervention. Patton (1996) argued that Scriven's formative/summative dichotomy takes a too narrowly defined view of evaluation by forcing all activity around evaluation into a formative/summative dichotomy and that evaluation is an open system that changes based on the purposes, values, and contexts of evaluators and their clients. In this way, Patton echoed Misanchuk in the layers or cyclic nature of evaluation and specifically introduces the concept of systems thinking in evaluation.

Moseley and Solomon (1997) continued the argument that confirmative evaluation creates a cycle of evaluation with formative and summative evaluation and made the evaluation processes more iterative and focused on continuous improvement. Moseley and Solomon also extended the explanation of confirmative evaluation by arguing that it uses multiple data-gathering instruments including self-reporting, work-sample analysis, performance analysis, context studies, and cost-benefit comparisons to ask if customers' expectations are being met by the products they are provided. For Moseley and Solomon (1997) confirmative evaluation

examines the value added by a product and asks if it should be continued, improved, or terminated. They argued that confirmative evaluation is necessary in continuous improvement organizations and acknowledged that the literature regarding confirmative evaluation at that time was insufficient and lacked empirical documentation.

Dessinger and Moseley (2015) significantly contributed to the study of confirmative evaluation with their *Confirmative Evaluation: Practical Strategies for Valuing Continuous Improvement*. This work further placed confirmative evaluation as a vital component of a system of evaluation and was the first writing with specific directions on conducting confirmative evaluations. In addition to providing steps and tools to perform confirmative evaluation, Dessinger and Moseley (2015) identified that confirmative evaluation is often not systematically conducted because it often occurs over long periods of time well after a training or intervention is implemented. As organizations and their needs change over time, original confirmative evaluation measures produced at the beginning of an intervention (if they were ever planned at all) often lack the flexibility and active monitoring to keep up with morphing organizational need. Dessinger and Moseley (2015) also acknowledged that confirmative evaluation seeks to establish the tangible and intangible merit or worth of a training process. This is a complicated endeavor that can often rely on extensive cost-benefit or return-on-investment analysis. Confirmative evaluation, simply put, is not easy. Unfortunately, Dessinger and Moseley's (2015) work is limited by the fact that it fails to fully demonstrate confirmative evaluation in action. Their single case study to demonstrate confirmative evaluation is of the Oz City Zoo and, while illustrative of how confirmative evaluation *could* work, is a hypothetical example in a fictional city, not an actual case of a real situation. Moseley and Solomon's (1997)

acknowledgement that the literature regarding confirmative evaluation lacked empirical documentation remained a significant gap in our understanding of confirmative evaluation.

After Dessinger and Moseley's (2015) work on confirmative evaluation, the concept of confirmative evaluation became a staple reference in HPT work related to evaluation. Moseley and Hasting's (2005) four-stage Intervention Implementation Process Model took a cyclical approach to continuous improvement that intentionally lays the groundwork for the use of confirmative evaluation in continuous improvement projects. Giberson et al. (2006) argued that confirmative evaluation can determine the extent of institutional change by demonstrating the results of a program and measures individual and organizational performance improvement and the results of change efforts. Additionally, Giberson et al. (2006) recognized that confirmative evaluation and summative evaluation are not distinct from each other so much by *when* they occur but by *why* they occur, and in their work confirmative and summative evaluation can be conducted simultaneously for different purposes. An evaluation's purpose informs decisions of what to evaluate and how and when to do so (Van Tiem et al., 2012).

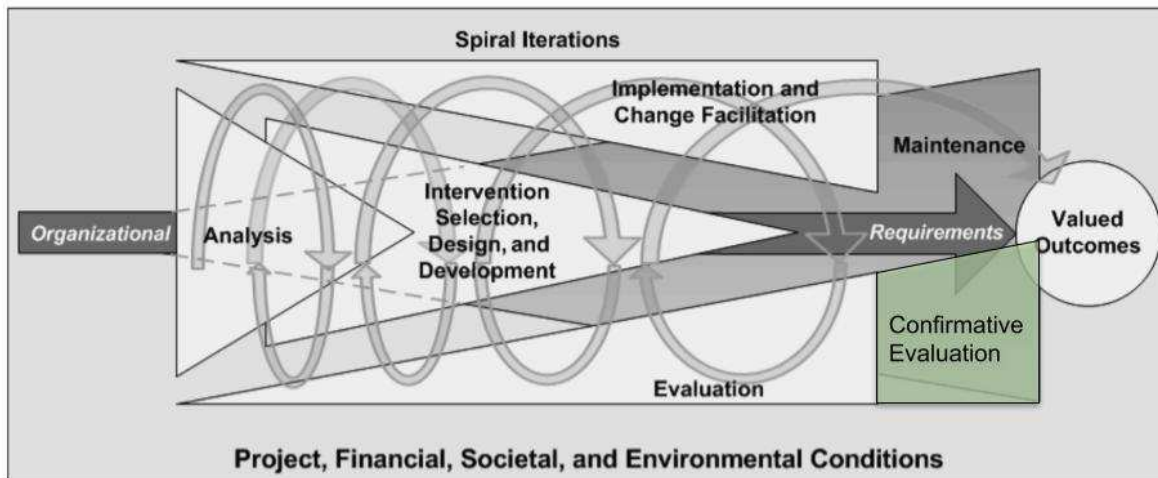
Van Tiem et al.'s (2012) HPT model further cemented confirmative evaluation's place in the literature by incorporating it into what became the de facto HPT model for ISPI. This model integrated Kirkpatrick and Kirkpatrick's (2006) four levels of evaluation and Phillips et al.'s (2006) ROI model into confirmative evaluation by identifying that confirmative evaluation examines an intervention's ability to sustain its effectiveness, efficiency, impact, and value. Further, Van Tiem et al. (2012) demonstrated the importance of integrating evaluation into the front-end analysis, intervention design and development, and the implementation and maintenance phases of an intervention. Their use of double ended arrows to connect evaluation to these three phases is meant to indicate the iterative nature of evaluation, highlighting the

importance of returning to evaluation frequently in a performance improvement project (Kang, 2012) and illustrates the model's adaptability to different situations and its flexibility to adjust to contextual constraints during an intervention (Van Tiem & Burns, 2020).

Marker et al.'s (2014) Spiral HPI Framework revises Van Tiem et al.'s (2012) HPT model to better illustrate the fluid, iterative nature of HPI work. Doing so, they argued, reflects the reality that HPI work is less linear than our models typically demonstrate. They additionally emphasize the difference between models and frameworks and chose the term *framework* to illustrate that performance improvement professionals use components of models that are relevant to their contexts. An issue in their framework visualization, and a demonstration of how confirmative evaluation is misunderstood generally, is that evaluation is cut short in the maintenance section. A truer representation (figure 2) of how evaluation should work, and where confirmative evaluation provides its greatest impact by determining the long-term impact and value provided by an innovation, is to have evaluation extend to the organization's valued outcomes. In this way, the maintenance of an innovation, the degree to which an innovation continues to meet organizational requirements, and the impact of an innovation on valued outcomes is clear. Confirmative evaluation is what can extend evaluation through maintenance and into the valued outcomes of an organization (figure 2). This revision also highlights the role confirmative evaluation plays in working after a change has been implemented to determine if the change has achieved its desired results.

Figure 2

Marker et al.'s (2014) Spiral HPI Framework, with Confirmative Evaluation



Challenges of Confirmative Evaluation

Since Misanchuk's original identification (1978c), confirmative evaluation has received growing but still insufficient attention, largely because of the relative difficulty in conducting confirmative evaluation and the challenge performance improvement professionals face in convincing their clients of the value of confirmative evaluation (Dessinger & Moseley, 2015; DeVaughn & Stefaniak, 2020a; Giberson et al., 2006; Zinoveff, 2008). Finding authentic examples of confirmative evaluation is additionally challenging because of the many barriers to evaluation generally. Marshall and Rossett (2014) identified that designers often do not evaluate, not because they lack the skill, but because of the many organizational barriers to that evaluation. Organizational barriers include environmental factors, lack of incentives, lack of tools and systems for evaluation and reporting, and lack of support from more skilled data analysts. Kennedy et al. (2014) echoed Marshall and Rossett's (2014) findings that level 3 & 4 of Kirkpatrick and Kirkpatrick's (2006) model, which evaluate the degree to which participants

apply what they learned and the degree to which desired outcomes occur, are realized as a result of the training and where confirmative evaluation generally focuses (Van Tiem et al., 2012; Zinoveff, 2008), are often not undertaken because of a lack of resources, organizational support, access to data, and trainers' lack of familiarity with evaluation methods to support level 3 & 4 evaluation. It is often overlooked because organizations often expend more energy on adopting a change than they do implementing it, and decisions are often made without follow up (Fullan, 1996), often because of the relative cost of conducting confirmative evaluation (Jackson, 1989; Guerra-Lopez, 2008).

In addition to organizational barriers to evaluation, there are additional internal barriers for designers and performance improvement professionals. DeVaughn and Stefaniak (2020b) found that evaluation is generally undertaught in IDT programs. They found that formative evaluation was most frequently taught in IDT programs and that summative and confirmative evaluation were often overlooked. They hypothesized that this may occur because of a lack of expertise among faculty or that this underteaching of summative and confirmative evaluation reflects the lack of emphasis from employers for these types of evaluation. A significant factor to this lack of understanding of confirmative evaluation as a field is the blurring boundaries of this unique form of evaluation. Because it shares similarities with both formative and summative evaluation, (Misanchuk, 1978a) confirmative evaluation is sometimes confused with these forms of evaluation that occur more commonly in the literature. For example, Kang (2012) assumed that summative and confirmative evaluation are essentially the same except for when they occur and since timing of evaluation is not a significant distinguishing factor, uses this argument to consolidate the reporting data in their study for both types of evaluation and revised ISPI 's HPT model to drop confirmative completely and replace meta evaluation with reflection. If timing

were the only difference between summative and confirmative evaluation, this conclusion would be logical, but it fails to acknowledge the difference in purpose and familiarity with the object being evaluated.

In another application of confirmative evaluation, Finney (2020) argued that confirmative evaluation could improve Stufflebeam's (1971) CIPP model. Stufflebeam (1971) argued that evaluation "is the process of delineating, obtaining, and providing useful information for judging decision alternatives" (p.267) and that the CIPP model divides decisions into contexts, inputs, processes, and products (Stufflebeam, 1971). Each class of decision, according to Stufflebeam (1971) comes with its own distinct form of evaluation. In Finney's approach, confirmative evaluation is synonymous with continuous improvement, and they place what they called confirmative evaluation immediately after each step of the CIPP model as a tie to the next step. While their study found that the remediation program had a positive impact on students, they failed to clearly explain how confirmative evaluation impacted the use of the CIPP model and actually mislabels formative evaluation as confirmative evaluation because it is being used during an implementation to ensure it is improved along the way. In addition to a timing issue in Finney's study, the more pressing gap in their study is the real heart of confirmative evaluation-- a decision of if the intervention should be continued, revised, or ended (Misanchuk ,1978a; Dessinger and Moseley, 2015). Scriven (1996) acknowledged the existence of gray areas between formative and summative evaluation, though they did not acknowledge this gray area could be confirmative evaluation and that the context and the purpose of evaluation dictated if it were formative or summative. This confusion of what is formative evaluation, summative, and confirmative evaluation highlights the dynamic nature of confirmative evaluation Misanchuk (1978a) originally identified and illustrates the difficulty of categorizing it. It also demonstrates

the lack of clarity in our field regarding different types of evaluation that must be addressed for us to better apply evaluation in our work.

A Typology of Evaluation

Because of this difficulty in categorizing confirmative evaluation, different ways of describing and conducting confirmative evaluation have developed, which in turn makes identifying and talking about confirmative evaluation even more difficult. Because evaluation has come to be correctly seen as a system of different types of evaluation, our field has fallen into the challenge of most systems views: it is impossible to identify and analyze all relevant aspects of a system because boundaries between elements of a system are vague (Adams et al., 2020; Giberson, 2015; von Bertalanffy, 1972). According to Giberson (2015), what a field of study needs to do is create boundaries within the system so it can develop deep knowledge about those areas. One step towards creating boundaries within a system is to define its elements. A typology of evaluation can clarify the similarities and differences among the many types of evaluation that make up a system of evaluation.

Typologies are useful so long as they clearly delineate among the forms of things they are meant to analyze. Scriven's (1996) formative/summative dichotomy admittedly leaves too much gray space to adequately define the different manifestations of evaluation. Scriven (1996) noted that their formative/summative dichotomy was simply one way to organize the different types of evaluation and that other typologies are useful for identifying and examining evaluation. Chen (1996) argued that valuative assessment is a necessary addition to Scriven's formative/summative dichotomy, but Chen's own four-way distinction lacks clear difference with Scriven's formative/summative dichotomy (Scriven, 1996). Zinoveff's (2008) typology attempts to delineate evaluation types by identifying seven types of evaluation in the literature:

Formative, Summative, Confirmative, Meta, Goal-based, Process-based, and outcome based. The problem with this typology is that there is not really an explanation of the differences between confirmative and goal, process, and outcome-based evaluations. They do make the brief argument that the 4th level in Kirkpatrick's evaluation model is really confirmative evaluation under a different name, but with the many types of evaluation presented in Zinoveff's typology, it is difficult to identify where one type ends and another begins.

Perez-Soltero et al. (2019) argued that evaluators have different needs when using evaluation methods and defined evaluation methods from which users can select a methodology based on the information desired, the type of evaluation, the timing of evaluation, and the tools used to conduct the evaluation. This work brings us closer to a more robust typology of evaluation, but it is inherently flawed because it relies on Zinoveff's (2008) 7 evaluation types, which confuses the relationship of goal-oriented, results based, and confirmative evaluations. Perez-Soltero et al. (2019) highlighted this confusion in their table of evaluation types. By identifying seven different types of evaluation and where they are applied in the literature, Perez-Soltero et al. (2019) noted the use of Goal-oriented, Process-based, and Results-based in 7 of 42 evaluation models in their typology. But in those seven models where evaluation is identified as results-based, process based, or goal-oriented, a strong case can be made that these are really forms of formative, summative, or confirmative evaluation. For example, Perez-Soltero et al. (2019) argue that the focus on organizational results and societal contribution makes Kaufman, Keller, and Watkins' evaluation model (Kaufman, 1996) results-based and goal-oriented evaluation. But when we recognize that confirmative evaluation's purpose is to determine the lasting impact and value of an intervention, we quickly see that organizational results and societal contribution are frames through which we can confirm an intervention's impact and

value, making results-based and goal-oriented evaluation examples or methods of confirmative evaluation. We see a similar issue in their argument that stakeholder-based evaluation is a results-based evaluation. Results-based evaluation could be considered a type of summative evaluation, especially considering that they note this type of evaluation to be conducted immediately after training where summative evaluation typically occurs (Dessinger & Moseley, 2015). Additionally, their analysis places the return on expectation of Kirkpatrick's (2009) Business Partnership Model as a results-based evaluation, but it fails to explain how an evaluation of the return on expectations is *not* confirmative evaluation while they identify levels 3 and 4 of the same model to be confirmative evaluation. We see similar issues in their classification of Draper's (2012), Griffin's (2012), Passmore and Valez's (2012) and Moldovan's (2016) models of evaluation. There is simply too much overlap in Perez-Soltero's (2019) typology to accurately organize evaluation based on the descriptors they selected.

Confirmative Evaluation by Another Name

To address the overlap issues in Perez-Soltero et al.'s (2019) typology, it is necessary to redraw the boundaries of the different types of evaluation. Doing so allows us to reconceptualize evaluation based first on its purpose rather than its output, timing, or tools used for the evaluation because these elements of any evaluation can be dynamic and change based on the purpose of the evaluation. When taken in this way, evaluation has four primary purposes: to improve a thing (formative), to determine if a thing fulfills the measure of its creation (summative), to determine the value or impact of a thing on its environment (confirmative), and to determine the degree to which effective evaluation has occurred (meta). Understanding an evaluation approach by its purpose first allows its other attributes like its timing, tools, and output to become clarifying details that help explain the evaluation being undertaken. This framework (table 2) for

considering evaluation intentionally returns to Dessinger and Moseley's (2015) full-scope evaluation model as the primary organizing principle and allows different methodologies to be more clearly understood as they relate to one another. This conceptualization allows for any number of descriptors to be identified in a particular model and acknowledges that an evaluation model will often reach into different evaluative purposes simultaneously while not requiring any model to be all inclusive. This flexibility in conceptualizing evaluation frees practitioners to apply those elements of evaluation that best fit their contextual constraints and expands their ability to consider evaluation methods that push the boundaries of previously defined models.

Table 1*Evaluation Conceptualization Framework*

		Models	
Evaluation Type <i>Purpose</i>	Descriptors	Four Level Evaluation Model (Kirkpatrick, 2006)	Training Outcome Evaluation Model (Moldovan, 2016)
Formative <i>To improve a thing</i>	Timing		
	Sources		
	Outputs		
	Etc.		
Summative <i>To determine if a thing fulfills the measure of its creation</i>	Timing	Immediately after	Before a training occurs, immediately after a training
	Sources	Surveys, interviews, assessments	Learner feedback, assessments
	Outputs	Level 1: Perception Level 2: Immediate Learning	Reaction evaluation, learning evaluation
	Etc.		
Confirmative <i>To determine the value or impact of a thing</i>	Timing	Up to a year after training	2-6 months after training
	Sources	Organizational data, observation, surveys, etc.	Trainee performance at workplace, Feedback from employer
	Outputs	Level 3: Trainees' performance improvement Level 4: Organizational performance improvement	Behavior evaluation, Results evaluation
	Etc.		
Meta <i>To determine the degree to which effective evaluation has occurred</i>	Timing		
	Sources		
	Outputs		
	Etc.		

Table 1 continued

		Models	
Evaluation Type <i>Purpose</i>	Descriptors	Kaufman and Keller Evaluation Model (1996)	Kirkpatrick Business Partnership Model (2009)
Formative <i>To improve a thing</i>	Timing		
	Sources		
	Outputs		
Summative <i>To determine if a thing fulfills the measure of its creation</i>	Timing	Immediately after; 2-6 months after training	Immediately after
	Sources	Unspecified	Interviews, questionnaires
	Outputs	Level 1: Perceptions of value and worth of process, methods and resources used Level 2: Learning Acquisition; Individual or small group payoffs Level 3: Behavior or performance; Individual or small group payoffs	Level 1: Trainee perception Level 2: Trainees immediate performance
Confirmative <i>To determine the value or impact of a thing</i>	Timing	more than 6 months after training	More than 90 days after training
	Sources	Unspecified	Interviews, Comparison of indicators, ROI metrics
	Outputs	Level 4: Results; Organizational Payoffs Level 5: Societal Contribution	Level 3: Improvements in trainees' performance Level 4: Improvements in organizational performance Level 5: Return on expectations
Meta <i>To determine the degree to which effective evaluation has occurred</i>	Timing		
	Sources		
	Outputs		

Table 1 continued

		Models	
Evaluation Type <i>Purpose</i>	Descriptors	Training Lifecycle Model (Griffin, 2012)	SOAP-M Evaluation Model (Passmore & Valez, 2012)
Formative <i>To improve a thing</i>	Timing	In development	
	Sources	Mixed methods	
	Outputs	Organizational and learner characteristics	
	Etc.		
Summative <i>To determine if a thing fulfills the measure of its creation</i>	Timing	Immediately after	Immediately after
	Sources	Mixed methods	Questionnaires, pre/posttests, psychometrics
	Outputs	Learning	Level 1: Self-assessment of learning Level 2: External assessment
	Etc.		
Confirmative <i>To determine the value or impact of a thing</i>	Timing	Over time	Over time
	Sources	Mixed methods	Observation, organizational benchmarks, profit, pre/post tests
	Outputs	Learning transfer, Impact, cost	Level 3: Achievements Level 4: Potential
	Etc.		
Meta <i>To determine the degree to which effective evaluation has occurred</i>	Timing		Long after training is complete
	Sources		Comparisons of many evaluations
	Outputs		Level 5: Meta-analysis
	Etc.		

This revision of how we view models of evaluation by focusing on their purpose first allows us to identify that confirmative evaluation, while not specifically mentioned in many evaluation models, has been a present but under-recognized component of these models. Our inability as a field to identify confirmative evaluation has left us unable to identify its value and explain it to our clients and those being trained in instructional design and performance improvement.

DeVaughn and Stefaniak (2020a) hypothesized that designers may use confirmative evaluation more if they had stronger models for it. Currently, our models do not clearly delineate how confirmative evaluation contributes to a performance improvement project. For example, Van Tiem et al.'s (2012) HPT model identifies that confirmative evaluation occurs, but it doesn't clearly identify when a practitioner switches from summative evaluation into confirmative evaluation or from confirmative evaluation back to formative evaluation (Hellebrandt & Russell, 1993; Moseley & Solomon, 1997). Most explanations of confirmative evaluation focus on the importance of timing, but they should instead emphasize when an evaluation's purpose changes. By doing so, we can see that the chronological boundaries that separate the types of evaluation are less important and fluid.

Another role of confirmative evaluation appears in Kalman's (2016) use of Marker et al.'s (2014) spiral HPI framework as a conceptual foundation to their hybrid evaluation-need assessment approach placed confirmative evaluation as a form of quality assurance and provides a model for incorporating confirmative evaluation as DeVaughn and Stefaniak (2020b) call for. For Kalman (2016), classifying a project as both evaluation and assessment affects its organization, framing, scope, purpose, research methods, and questions while increasing study efficiency and adaptability. This increased efficiency should make conducting confirmative

evaluation a stronger possibility for performance improvement professionals as it provides a stronger tie to continuous improvement and can help make the case for continued intervention development. But this approach highlights the challenge that the formative/summative/confirmative boundaries blur together in reality, making it difficult to identify where one form of evaluation ends and another begins. Kalman (2016) called for additional case studies that explore the cyclical nature of the relationship between evaluation and needs assessment, but we've not yet seen these studies in the literature.

Conclusion

Binder (2001) argued that measurement in the field of performance improvement occurs for three reasons: validation, accountability, and decision making. Confirmative evaluation can meet all three of these reasons, but we have had a difficult time in our field helping stakeholders see the value confirmative evaluation provides for validation, accountability, and decision making. What is needed at this point in our field are solid examples of confirmative evaluation beyond hypothetical situations in mythical cities (Moseley & Solomon, 1997; Dessinger & Moseley, 2015). As we develop these examples, we will see that the distinguishing feature of confirmative evaluation is its purpose: to determine the impact or value of a thing. By reconceptualizing evaluation models based on their ability to improve a thing, to determine if a thing fulfills the measure of its creation, to determine the impact or value of that thing, or to determine the degree to which effective evaluation has occurred, we see that we already have many useful models for conducting confirmative evaluation in the current human performance literature. Still, organizational and designer-specific constraints limit our ability to conduct evaluations. To address this limitation, designers and performance improvement professionals should adopt evaluation methods that are relevant to their context and their skill sets (Van Tiem

and Burns, 2020). By reaching beyond instructional design literature into other fields that rely on evaluation, designers can find methods that work for the organizations they seek to improve.

The following chapter will outline one way confirmative evaluation can be conducted by adapting an industrial production evaluation technique for an online higher educational context that serves organizational need and confirms the relative impact of a structural change to the organization. The results of this confirmative evaluation were a stronger understanding of how well the new organizational model is functioning after its implementation (validation), insight into what elements of the change have not yet been fully implemented (accountability), and clear areas of focus for future revisions and innovations (decision-making). For the Performance Improvement field, it serves as an example of the role confirmative evaluation can play in the change management process and how confirmative evaluation adapts to contextual needs by adopting unique methodologies to achieve organizationally significant goals while helping fill the gap we have of empirical examples that will help us understand confirmative evaluation.

CHAPTER III

METHODOLOGY

The purpose of this instrumental case study was to develop a deeper understanding of confirmative evaluation and to help fill the gap in the literature of examples of confirmative evaluation. In this case, confirmative evaluation was employed in online higher education administration by determining the degree to which an organizational change has been institutionalized, identifying the effects of that change, and noting where the change needs further reinforcement. The questions in this study were

1. How can confirmative evaluation be used to determine the degree to which organizational changes have been institutionalized?
2. To what degree can confirmative evaluation determine the effects of changes in continuous improvement efforts?
3. How does data from confirmative evaluation support continuous improvement efforts to reinforce organizational changes?

Instrumental Case Study Design

Instrumental case study can be used to provide insight in an issue or to refine theory (Baskarada, 2014). Instrumental case studies differ from intrinsic case studies in that instrumental case studies use a case to understand a larger phenomenon by observing it within a case (Harling, 2002; Mills et al., 2010). Unlike intrinsic case study, it focuses less on the complexity of the case and more on the specifics related to its research questions (Mills et al., 2010). Because of this, specific questions are more common in instrumental case studies compared with more open-ended questions used in intrinsic case studies. (Mills et al., 2010). Instrumental case studies are conducted to provide insight into a larger issue or redraw a

generalization (Stake, 2003). While the case study is valuable to understand the case it studies, its primary role is to illuminate a larger phenomenon (Stake, 2003). Instrumental case study is conducted by offering thick description of the case and relies on triangulation of data to support conclusions drawn during the case (Mills et al., 2010). While instrumental case study doesn't guarantee generalizability beyond the case, it does identify patterns and themes that can be compared with other cases, providing a comparative point for other cases in which the same phenomenon might appear (Mills et al., 2010).

Identification of the Case

Institutional Context

The university in this study had an Online Learning department that employs approximately 2,000 online adjunct instructors to serve over 60,000 students enrolled in online courses at three institutions in a higher education system. Starting in the Winter 2021 semester and fully enacted in the Spring 2021 semester, this Online Learning department underwent an organizational restructuring that simplified instructor teams and reporting lines. While this change brought many benefits of simplicity and cost savings, it worked against a decade of organizational philosophy, policies and procedures, and organization-specific software development.

The Online Learning department of the university in this study had enacted changes to the organizational structure of online adjuncts for three reasons. First, Online Learning sought to better align online adjuncts with the academic departments for which they teach. The hope in doing so was to create a stronger connection and working relationship between departments and online adjuncts to foster more collaboration that would lead to faster course improvements and to create a sense of collegiality that would help online instructors feel greater belonging and sense

of purpose in their work with the university. Second, Online Learning sought to align online instructors' focus with their primary responsibility of teaching their courses. While ancillary activities like professional development and communication among online adjuncts are valuable, the old organizational model created a level of focus on these activities that distracted online instructors from their primary purpose of engaging in their courses. Third, Online Learning sought to streamline its leadership model to increase system efficiencies and to reduce administrative cost.

Because this restructuring works against so much organizational history and context, the migration to the new organizational model required substantial planning, messaging, training, and change management to ensure all stakeholders understood the changes, how those changes affected them, and their new roles and responsibilities. With the start of the Fall 2021 semester, all online courses had been in the new organizational model for over four months and the university's Online Learning department was at a stage in this metamorphosis where confirmative evaluation could help determine if the changes enacted thus far had resulted in the desired changes in remote leader activity.

The change in how online instructors were organized and managed was a significant organizational shift that upended a decade of organizational policy and the institutional culture that developed around it. At the start of the Fall 2021 semester, the new organizational model had been in place as the sole organizing structure in Online Learning for enough time to determine if the change has taken root (Dessinger & Moseley, 2015). Confirmative evaluation serves the purpose of determining the impact of a thing, in this case an organizational restructuring and job revision, and extends traditional evaluation measures to an organization's valued outcomes (Marker et al., 2014). It also links prior interventions to future improvement

projects in an iterative process (Hellebrandt & Russell, 1993; Moseley and Solomon, 1997) by identifying further needed changes (Moseley & Hasting, 2005; Giberson et al., 2006) and the degree to which changes have been institutionalized by an organization (Giberson, et al., 2006). The Fall 2021 semester was the ideal period in which to determine if the enacted organizational changes had the desired effects and to continue to improve the new organizational model. This project would help confirm the degree to which the organizational restructuring and revised leader duties has met its intended goals and was creating sufficient value by examining how remote adjunct leaders spent their time in the semester.

To this end, this case study intended to confirm how well the organization abandoned the old organizational model of in favor of the organizational model by examining the activity of remote leaders in the new organizational model and how that work aligned with their defined roles found in their respective job descriptions. In the old model, remote leaders' primary function was to provide support and mentoring for other online adjunct instructors. Under the new organizational model, remote leaders were meant to focus their efforts on course design and delivery improvement. If remote leaders had continued to function as they had before the organizational restructuring, Online Learning will have done little more than put old wine in new bottles and failed in its efforts to change the focus of remote leaders' primary work and subsequently the nature of the organization at large.

Central to the success of the new organizational model was the work remote leaders do in their contracted responsibilities. This model was led by two types of remote leaders: Online Course Representative (OCR) and Assistant Course Leads (ACL). Their work and the time allotted to that work was substantially different from the old remote leaders' duties and time, and the Online Learning department needed to be certain both types of remote leaders were

performing their jobs as they were designed. While the OCR and ACL roles were extremely similar, they were different in that ACLs had authority OCRs do not to make fixes in the master course. This difference, along with the different sizes of course groups ACLs and OCRs each served, made each remote leader role unique. Understanding the similarities and differences of these roles as they were deployed could help the organization determine where each role could be best deployed within the organization. Additionally, Online Learning needed to ensure that the designed job and contract load adequately fit the ACL and OCR roles. Accurate job descriptions provide value to organizations in terms of employee satisfaction, job performance, stress, and accountability (Ashraf, 2017; Mohamed & Hossny, 2020; Pató, 2017). In the context of this case study, the job descriptions of ACLs and OCRs were the foundation for determining the degree to which the new remote leadership roles were being implemented.

If the job as it was designed required too much time of OCRs and ACLs, they would likely suffer from burnout and exhibit poor performance. Likewise, if ACLs and OCRs went beyond what was expected of them, they could create unnecessary dependencies among the instructors they served and distract ACLs and OCRs from their other vital job duties. This study helped determine if the desired changes had been effectively implemented and aligned with the remote leaders' job descriptions. From this information, the organization's next steps in cementing the new organizational model and needed revisions to the remote leader job descriptions and trainings would become clear.

Study Population

Participants in this study were 342 of the 363 Assistant Course Leads and Online Course Representatives employed by this university during the Fall 2021 semester. Assistant Course

Leads and Online Course Representatives were online adjunct instructors that were representative of the total population of online instructors at this university (see

Table 2) All participants had previously completed the required trainings to serve in these positions prior to their service as ACL or OCR and are assumedly well trained for their job tasks.

Table 2

Online Instructor and Course Group Leader Demographics

	Online Instructor Population (%)	ACL and OCR (%)
Gender		
Female	873 (45%)	156 (43%)
Male	1048 (55%)	207 (57%)
Highest Degree		
Bachelors	277 (14%)	52 (14%)
Masters	1429 (74%)	268 (74%)
Doctorate	207 (11%)	38 (10%)
Unspecified	8 (0%)	5 (1%)
Average Semesters Taught	12	13
Geographic Region-US		
Northeast	20 (1%)	5 (1%)
Southeast	264 (14%)	54 (15%)

Southwest	174 (9%)	30 (8%)
Intermountain West	1178 (61%)	214 (59%)
Midwest	76 (4%)	21 (6%)
Pacific Coast	207 (11%)	35 (10%)

Participant Selection.

Because there were only a few course groups with more than 21 instructors (see Table 3), leaders of those groups were excluded from this study. The remaining 351 remote leaders were invited to participate via an email explaining the research project and its impact. Nine remote leaders opted out of the time study and were excluded from data collection. The remaining 342 remote leaders were sent the time study data collection tools and reminder emails throughout the Fall 2021 semester.

Table 3

Course Group Size

Size of Course Group	Number of Groups led by OCR (%)	Number of Groups led by ACL (%)
Micro (1-5)	188 (52%)	90 (25%)
Small (6-20)	43 (12%)	30 (8%)
Medium (21-40)	2 (1%)	6 (2%)
Large (41-60)	1 (0%)	2 (1%)
Mega (61+)	0 (0%)	1 (0%)
Total	234 (65%)	130 (36%)

Role of the Researcher, Relationship with Participants, and Protections for Participants

The researcher in this study served as an Instruction Manager within Online Learning, and as such held a position of authority over the participants in this study as study participants reported indirectly to the Instruction Manager. Such relationships had the potential for creating coercive biases in the data collection process. To mitigate this issue, the researcher ensured all participant responses were collected anonymously using survey software that would not identify personal information that would allow linking a particular response to an individual participant. Further, because course groups that were larger than 21 instructors were relatively rare in the organization, leaders of course groups with more than 21 instructors were not included in this study as it would be nearly impossible to guarantee their anonymity (see Table 3). Participants also retained complete control over the time study tracking sheet that was offered to help them keep track of their data. Finally, a matrix sampling methodology was employed to facilitate data collection and to increase protection for participants if anonymity should be compromised by intentionally limiting the self-reported data from any single participant.

Sources of Evidence

To address the research questions guiding this case and to support triangulation of conclusions drawn in this case study (Mills et al., 2010), this study used three sources of evidence: (a) Remote Leader Time Study with matrix sampling; (b) Course Group Survey; and (c) Course Council Survey. Each source of evidence is described below.

Remote Leader Time Study

The remote leader time study examined the time spent by OCRs and ACL at this organization in their respective roles as remote leaders. Time study methodology is common in manufacturing and industrialized industries and functions by dividing processes or work into its

sub-processes and actors, evaluating the behavior of the sub-processes and actors, and summing all parts of the process to obtain an overall system timing (Yazdi et al., 2019). Work measurement studies often stem from Taylor's (1911) work on scientific measurement to standardize work and improve productivity. Time study is a form of work measurement, and can benefit managers, workers, and consumers by reducing labor costs and providing performance standards to support managerial decisions, if operators to be studied are fully trained to perform the work with average skill and effort (Karger & Bayha, 1987). Time studies can identify the standard amount of time needed to perform a job and allows for comparison of actual performance against job design expectations following a ten-step process (Kulkarni et al., 2014).

Engineered, manufacturing, or industrialized work measurement methods do not always translate to knowledge worker or service worker situations because of the inherent complexity of knowledge and service work and the many means by which knowledge workers can approach the same task (Pepitone, 2002). That said, time studies can be used to identify the number of instances of particular tasks knowledge or service workers perform to calculate the normal time it takes to perform a job and can serve as the basis for measures of knowledge and service workers' efficiency and productivity (Klassen et al., 1998). Benefits of measuring knowledge worker productivity include improved personnel selection, job design, identification of redundancies, strategic planning, gap analysis, and establishing performance benchmarks (Ramírez & Nembhard, 2004). While there is no generalized method to measure knowledge workers' productivity because their work is intangible and difficult to categorize (Ramírez & Nembhard, 2004), the basic approach of a time study is to divide a job to be studied into its component acts, observe workers doing these acts, and calculate the time it takes a standard,

trained worker to complete each act with necessary breaks and interruption allowances (Kulkarni et al, 2014).

To answer the research questions in this case, a time study over the course of a typical semester was used to track the time spent by remote leaders (ACLs and OCRs) on a weekly basis and to compare the results of this study with job-defined activity expectations. The weekly nature of this study revealed the ebb and flow of time required of course group leaders throughout the semester and revealed how time spent on different areas of responsibility increase and decrease over the course of a semester. The results of this study confirmed the degree to which online remote leaders at this institution were performing their work as it was designed, and it could additionally inform revisions to the job design and training of remote leaders.

Matrix Sampling

A time study such as this could face considerable challenges like ensuring an adequate number of respondents complete the entire survey used to collect the needed data over the entire semester or producing too much data to compute efficiently (Gonzales & Eltinge, 2007; Thomas et al., 2006). The study required collecting data from participants each week for fifteen weeks. Because of the extended nature of this survey, requiring a sample of participants to complete the entire survey would prove impractical, and participants would likely experience survey fatigue, thereby diminishing the quality and quantity of reported data. Methods to scale down the survey and data analysis processes included reducing the number of weeks in which data will be collected, but this would reduce the efficacy of the entire study by creating holes in the semester view of OCR and ACL activity.

A better approach in this situation was to employ a matrix sampling methodology (Thorndike & Thorndike-Christ, 2010) to track course group leader activity throughout the Fall

2021 semester. Matrix sampling divides especially long surveys into smaller, more manageable surveys applied to groups of respondents when too long of surveys lead to decreased response rate, when collecting all the data from individual respondents would take too much time, or when there would be too much data to efficiently compute (Gonzales & Eltinge, 2007; Thomas et al., 2006). This study was unique because most matrix sampling efforts split long surveys of different questions among many respondents while this study used a short survey that would be applied many times over many months among many respondents. Under this conceptualization, this study was a survey of the different activities required of remote leaders over each week of the semester, and the weeks of the semester were the grouping logic that organized the sampling. While matrix sampling is not useful for decisions about individual performance, it is highly useful for understanding how an entire group performs (Thorndike & Thorndike-Christ, 2010). Matrix sampling was an appropriate methodology for this project because it provided a means to create a week-by-week view of the remote leadership roles in this study while not overtaxing a sample of participants or leaving gaps in the knowledge of the entire semester contract for remote leaders.

Additionally, matrix sampling provided a means to minimize the potential perceived risk for any single member of the study and to reduce administrative cost (Childs & Jaciw, 2002; Kaplan & Su, 2016). By involving all remote leaders and collecting their responses anonymously, individual responses could not be tracked to individual leaders. Additionally, remote leaders in this study only reported on their activity a limited number of times in the semester, so the data on any single remote leader was intentionally incomplete. This method reduced the amount of potentially self-incriminating evidence a remote leader would be asked to report, and if participant anonymity were breached, collected data of a single individual was

intentionally insubstantial to warrant punitive measures. Matrix sampling, then, provided a measure of protection for participants while providing a means to collect sufficient data to understand the work of online remote leaders at this university.

Extant Data

This case also relied on extant organizational data collected through tools previously designed by the organization and explained below.

Course Group Survey

In this case, a course group is comprised of all online instructors teaching a specific course. For example, all Spanish 101 online instructors were organized into a course group, led by a remote leader (ACL or OCR) who also taught Spanish 101. All online instructors had previously been organized into course groups across the university. The Course Group Survey was a tool developed by the university to collect observational data from the instructors in the course group each semester and asked group members to identify what activity they observed of their remote leaders throughout the semester. Survey data from Spring 2021 and Fall 2021 were used in this case study. Winter 2021 data was unavailable because the course group survey was not administered in that semester.

Course Council Survey

In addition to having a course group, every course is assigned a course council, which consists of an on-campus faculty member responsible for the course (Course Lead), a curriculum designer assigned to the course, and the remote leader (OCR or ACL) of the course group. The Course Council Survey was a tool previously developed by the university to identify satisfaction with remote leaders' work from the course lead and curriculum designer as they served with the

OCR or ACL on the course council. Survey data of course leads from Fall 2021 were used in this case study to triangulate results of the time study.

Instruments

The time study in this case required accurate tracking and reporting of participant time across time and space over the Fall 2021 semester. To accomplish this, several tools were used.

Time Study Sheet

The Time Study Sheet was a spreadsheet only used by participants for tracking purposes. It identified the remote leader's job tasks and allowed the participant to keep track of each activity daily for one week. The spreadsheet automatically summed each category for a total of each activity over the week, making it easier for participants to accurately track their total work for the assigned week. All data entered on the spreadsheet was only accessible by that participant and was used as a tool to help the remote leaders accurately track their activity over a single week.

Time Study Survey

After collecting data for one week using the Time Study Sheet, participants were emailed a link to an anonymous survey where they reported the totaled amounts for their respective week. The survey also asked for information regarding the size of course group to allow for sorting responses.

Both the Time Study Sheet and Time Study Survey were based on tools used in previous time study work in the organization. During the Fall 2020 semester, other remote leaders participated in a similar time study to evaluate the time they spent in their typical work. In this

internal study, participants reported their activity in two weeks of the semester using a time study sheet and anonymous survey. Prior to that study, a pilot study was conducted during a six-week summer term with remote leaders. All remote leaders of that summer term ($n = 12$) completed the survey. Individual scores were compiled for mean activity for each question. Total time spent by remote leaders over 14 weeks on both surveys was approximately two hours. These findings suggest the remote leaders were able to adequately use the tool to accurately track their time in the week. When asked about the data collection process, remote leaders indicated the process was relatively easily, though they indicated remembering to track their time proved difficult. This was remedied in the Fall 2020 semester by sending participants reminder emails throughout the week and resulted in an 81% survey completion rate. The results of this previous study showed the average amount of time remote leaders spent in their job duties and how the time they spent in their different job duties shifted over the course of the semester and served as a benchmark for the current and future time studies for the organization.

Time Study Data Collection Procedures

In the time study for this case, remote leaders were divided by size of course group into five categories based on current pay structures and size parameters for course groups (see Table 3). I then randomly divided course group leaders into 15 survey groups by drawing on course group leaders from the micro and small groups. These 15 survey groups corresponded with the fourteen weeks of the regular semester plus the week before the semester begins (T-1) when instructors begin preparing their courses. Each course group leader was invited to participate in data collection and sent the anonymous survey four times in the semester. Leaders of medium (21-41 instructors) large (41-60 instructors), and mega (more than 61 instructors) course groups, were not reported in this study as their anonymity could not be protected in the data collection

process. As these groups represent 3% of total course groups, their exclusion from this study does not represent a significant gap in our understanding of how course groups function, but it does create a limitation regarding the applicability of this research to larger course groups. Large and Mega course groups could be studied using different methods but lie outside the scope of this project simply for project manageability.

The Friday before each collection period, each course group leader received an email informing them of the assignment to report on their time spent in the upcoming week. The email included the Time Study Sheet participants could use to track their activities during that week. The Time Study Sheet automatically totaled each activity over the week, and participants submitted the results in their total column on the Monday following the week of data collection using an anonymous Qualtrics survey that corresponded with the categories on the worksheet. To encourage regular tracking of activity and to discourage estimating when they complete the survey at the end of the data collection period, participants received reminder emails on the Wednesday and Friday of each week they collected data.

Analysis

Research Question 1 is *how can confirmative evaluation be used to identify the degree to which organizational changes have been institutionalized?* To answer this question, I compared the results of the time study with the estimated time stated in the job descriptions for the remote leaders' roles. I averaged the total amount of time spent by remote leaders in their different duties based on the size of course group and type of leader contract. Those items were then summed as they apply to the principal duties identified in the remote leaders' contracts. I calculated the amount of time remote leaders are expected to spend in each category based on the leader role and the group size and found the difference between average reported time and

expected time to find the job discrepancy (Barbouletos, 2011) for remote leaders. The difference between average time and expected time demonstrated the degree to which remote leaders are performing the job as it was defined and how well the different roles had been institutionalized.

Additionally, I compared these results with the data from the Fall 2021 Course Group Survey and the Fall 2021 Course Council Survey. The Course Group Survey collected data from course group members served by the remote leader and identified the activities course group members observed the remote leader performing in their duties. I averaged the number of reported instances of the different activities by remote leader type and group size in the time study and looked for alignment between the time spent and the number of instances of specific activities observed by group members. The alignment between observed activity and reported time confirmed the degree to which remote leader activity was noticed by those they served and suggested the degree to which the role was being performed as it had been designed.

The Course Council Survey collected course lead observation of remote leaders' work. These results were triangulated with the time study and the course group survey to further check for fit of the time study data and to observe any alignment that arose in the data. In relation to RQ1, this confirmed the accuracy of the time study report and further helped identify to what degree the new job duties had been institutionalized.

Research Question 2 is *to what degree can confirmative evaluation identify the effects of changes in continuous improvement efforts?* To answer this question, I searched for improved course council and course group satisfaction with the remote leaders' performance. Data from Winter 2021 & Fall 2021 Course Council Surveys revealed Course Council satisfaction across semesters after the changes have been implemented. The difference between Winter 2021 & Fall 2021 data revealed net changes on item level and total satisfaction. Likewise, the difference

between Winter 2021 & Fall 2021 Course Group Survey data revealed net change on item level and total satisfaction. A positive difference in data from either study suggested growing Course Group satisfaction and subsequent value from the new model, negative difference suggested decreasing course group satisfaction and subsequent decreased value from the new model, and neutral difference suggested no change in satisfaction or subsequent value as it applied to course group satisfaction.

Research Question 3 is *how does data from confirmative evaluation support continuous improvement efforts to reinforce organizational changes?* To address this research question, I used the results from RQ1 and RQ 2 to identify gaps in remote leader performance and their potential root causes. After identifying the average time spent by remote leaders based on the type of group and group size and determining the degree to which course councils and course groups were satisfied with remote leader performance, I compared results from RQ1 and RQ2 to identify discrepancies between course group and council overall satisfaction and remote leaders' reported time. Incongruence between the time study and course group observations suggested a disconnect between remote leaders and those they serve in the course group, identifying another gap in the adoption of the new leadership model.

As the next step in the confirmative evaluation process, I then identified potential causes for these gaps. To establish boundaries around my case study, I intentionally ended my study before analyzing potential causes further or making recommendations for their remedy as that effectively begins the next round of the continuous improvement process.

Limitations

An instrumental case study like this comes with inherent limitations. As noted above, leaders of medium, (21-40 instructors), large (41-60 instructors) and mega (more than 61

instructors) course groups, were not reported in this study as their anonymity could not be protected in the data collection process. As these groups collectively represented 3% of total course groups, their exclusion from this study did not represent a significant gap in the understanding of how course groups function, but it did create a limitation regarding the applicability of this research to larger course groups. If medium, large, and mega course groups become more common in the future, additional time studies including these groups may be warranted.

Additionally, the data is intentionally not compared by academic department or college, nor does it include the administrative cost of the remote leader program. These analyses would also prove fruitful and provide a more thorough valuation of each program but doing so required substantial organizational analysis that would extend beyond the scope of this project and would have significantly increased the complexity of the required analysis and extend the purpose of the study beyond its stated intentions. Time studies are inherently complex endeavors, and increasing the complexity generally decreases the study's feasibility. Future analysis of existing data after this study could be warranted to explore these areas.

Another limitation in this study was its reliance on self-reported activity. Optimal time study methodology would have an observer track time of workers in their job (Overby, 1983), but as remote leader work in this institution is performed at different times and in different geographic locations, direct observation was impossible. In addition, all course group leaders were adjunct faculty with no guarantee for continued employment from one semester to another, so there existed the potential for course group leaders to feel pressured to report more work in each week than they actually performed to ensure their good standing and to guarantee ongoing contracts. Messaging was included repeatedly throughout the study reminding course group

leaders that their reports were completely anonymous, that the data collected would only be reported in aggregate, and that their accurate representation of the work they do would improve the design of their job in the future to mitigate this potential for data fixing. These measures, along with sufficiently large data pools and analytic statistics, helped mitigate this limitation.

Conclusion

Even with these limitations, this instrumental case study was extremely beneficial. The university in this study benefited from this project as it learned how well the remote leaders aligned with current job description and gathered ways to improve that alignment. Further, as a work of confirmative evaluation, this study reveals to the university the comparative value of the new organizational model against the previous model. The university could also use this research to evaluate its change management approach to this substantive change in its online adjunct organization and the data collected here helped build the university's knowledge and understanding of how its total online learning program functioned as a system.

Regarding the field of performance improvement, confirmative evaluation is a vital but often overlooked component of evaluation that can be used along with formative and summative evaluation (Giberson et al. 2006). As an instrumental case study, this study showed practitioners how confirmative evaluation can be conducted and provides a real-world example, helping fill a gap in our literature around confirmative evaluation. This study also demonstrated how confirmative evaluation is an integral part of continuous improvement efforts by determining the degree to which changes have been implemented and identifying where further changes may be warranted for the organization.

Finally, part of the challenge of confirmative evaluation is its relative cost for evaluating decisions that have already been made (Guerra-Lopez, 2008; Jackson, 1989), and this study demonstrated that the cost of confirmative evaluation can be mitigated through combining new research with extant data and can provide a deeper analysis of a training program or intervention's total impact, thereby leading to further innovation and performance improvement.

CHAPTER IV

RESEARCH FINDINGS

Instrumental case studies are useful for refining theory (Baskarada, 2014) and serve to help a field understand a larger phenomenon (Harling, 2002; Mills et al., 2010). As an instrumental case study (Baskarada, 2014) aimed at exploring how confirmative evaluation can be conducted by performance improvement professionals in continuous improvement efforts, this study filled a gap in current literature around confirmative evaluation by presenting a real-life example of how confirmative evaluation works in an organization. This instrumental case study can serve as a model for other designers as they consider how confirmative evaluation can be conducted in unique contexts (DeVaughn and Stefaniak, 2020a) while demystifying confirmative evaluation by drawing clearer boundaries among confirmative, formative, and summative evaluation (Misanchuk, 1978a, Dessinger & Moseley, 2015; Giberson, 2015).

In this case, an online learning department that served three institutions in a private higher education system had undergone significant organizational restructuring around its management and organization of online adjunct instructors. Changes were fully implemented in Spring 2021, and in the Fall 2021 semester, a confirmative evaluation project was undertaken to determine the degree to which some elements of these changes had been implemented. The focus of this case were the online remote leaders (Online Course Representatives and Assistant Course Leads) used to lead groups of online adjunct instructors (course groups) and coordinate with course councils (course leads and curriculum designers). This case sought to determine the degree to which OCRs and ACLs performed their recently redesigned duties according to their new job descriptions and the effects of their work within these new job duties on course groups and course leads.

Within this context, the questions in this instrumental case study about confirmative evaluation were

1. How can confirmative evaluation be used to determine the degree to which organizational changes have been institutionalized?
2. To what degree can confirmative evaluation determine the effects of changes in continuous improvement efforts?
3. How does data from confirmative evaluation support continuous improvement efforts to reinforce organizational changes?

These questions were addressed in the case through a time study of remote leaders' work during the Fall 2021 semester and a comparison of the results of that study with the online learning department's existing surveys to evaluate the satisfaction of course groups and course councils with the work of remote leaders.

This chapter outlines the findings from this case as they address the research questions presented above. It then summarizes the results in preparation for the final discussion and summary in the following chapter.

Research Question 1

Research Question 1 was *how can confirmative evaluation be used to identify the degree to which organizational changes have been institutionalized?* To answer this question, I conducted the remote leader time study and compared the results of the time study with the estimated time stated in the job descriptions for the remote leaders' roles. I then triangulated these results with the results of the organization's Course Group Survey and Course Council Survey to confirm the accuracy of the time study results and to draw further conclusions about the study's results.

Time Study Overview

Course groups at this organization were divided by group size (see Table 3). All micro and small group ACLs and OCRs were invited to participate in the time study (N=351) while medium, large, and mega group remote leaders were excluded because their anonymity could not be guaranteed in the study. Micro and Small group remote leaders were informed via email of the purpose of the study and its impact. Nine remote leaders opted out of the time study and were excluded from data collection. The remaining 342 remote leaders were sent the time study data collection tools and reminder emails throughout the Fall 2021 semester. Each week, approximately 92 remote leaders were invited to participate in the time study. They were asked to keep track of their time in their remote leader duties using the provided Time Study Sheet and then sent a link on the Monday of the following week to the Time Study Survey to report their activity for the previous week. The time study survey received an average of 33.93 responses weekly, representing a 37% response rate (See Table 4).

Table 4

Time Study Response Rate

Week	Sent	N	Response Rate	Week	Sent	N	Response Rate
1	95	37	0.39	8	94	42	0.45
2	94	24	0.26	9	94	37	0.39
3	88	34	0.39	10	89	29	0.33
4	93	35	0.38	11	91	35	0.38
5	96	43	0.45	12	95	36	0.38
6	90	47	0.41	13	91	31	0.34
7	89	28	0.31	14	89	27	0.30
				Total	1288	475	0.37

At the end of the Fall 2021 semester, I compiled the data and averaged the total amount of time spent by remote leaders in their different duties based on the size of course group and type of leader contract. Those items were then summed as they apply to the principal duties identified in the remote leaders' contracts. I calculated the amount of time remote leaders are expected to spend in each category based on the leader role and the group size and found the difference between average reported time and expected time to find the job discrepancy (Barbouletos, 2011) for remote leaders. The difference between average time and expected time can be represented as $AWT - EWT = DT$ and demonstrated the degree to which remote leaders were performing the job as it has been defined and how well the different roles had been institutionalized (see Table 5). When calculated using the formula above, the data demonstrated that on average Assistant Course Leads of Micro groups spent 32 minutes less time than contracted and ACLs of Small groups spent 15 minutes less time weekly on their duties than contracted. Online Course Representatives leading Micro groups spent 3 minutes more time on average each week, while OCRs of Small groups spent 2 minutes less on their duties than contracted each week. The average weekly time of all groups combined was 11 minutes less than expected. High standard deviations for all groups show that remote leader activity within each group varied widely.

Table 5

Comparison of Remote Leader Role & Size Weekly Average

Role & Group Size	Expected Weekly Time (Minutes)	N	Average Weekly Time (Minutes)	Difference in Time (Minutes)	Difference in Time %	SD
ACL Micro	120	123	87.83	-32.17	0.73	114.87
ACL Small	180	68	164.97	-15.03	0.92	181.35
OCR Micro	60	237	62.70	2.7	1.05	81.83
OCR Small	120	46	117.80	-2.2	0.98	132.37
Total	120	474	108.33	-11.43	0.92	127.61

Comparison of Remote Leaders by Institution

Online Learning in this system serves three different institutions. Two of these three institutions participated in the time study. Institution I is a four-year university in the intermountain west, and Institution E is a two-year college also located in the intermountain west. The Instructor Manager that served Institution E asked specifically for a comparison of remote leaders at Institution E with the remote leaders at Institution I because course design and maintenance processes that may impact remote leader activity differ across these institutions. When comparing data across institutions, difference in time was observed among remote leaders (see Table 6). Online Course Representatives leading micro groups at Institution I (N=236) averaged spending three minutes more than their contracted time while OCRs at Institution E (N=1) met their contracted time perfectly, but with such a small sample of micro group OCRs at Institution E, little can be drawn from this data. As there were no OCRs at Institution E leading small groups, there is no comparative data for this set. Micro group ACLs at Institution I

(N=104) spent 30 minutes less than their contracted time while ACLs as Institution E leading micro groups (N=19) spent 44 minutes less than their contracted time. Finally, the discrepancy between ACLS leading small groups at Institution I (N=54) and Institution E (N=14) was the greatest with small group ACLs at Institution I going 3 minutes under their weekly contract and small group ACLs at Institution E going 61 minutes under their contracted time. Standard deviations for all groups showed significant variance among remote leaders' time overall, and average time of ACLs compared by institution showed a difference in time required.

Table 6

Comparison of Time Spent by Role, Institution, and Group Size

Role	Institution	Group Size	N	Contracted time	Average time	Difference in time	Difference in time %	SD
OCR	I	Micro	236	60	62.71	2.71	1.05	82.00
OCR	E	Micro	1	60	60	0	1.0	NA
OCR	I	Small	46	120	117.80	-2.2	0.98	132.37
OCR	E	Small	NA	120	NA	NA	NA	NA
ACL	I	Micro	104	120	90.06	-29.94	0.75	122.24
ACL	E	Micro	19	120	75.63	-44.37	0.63	61.12
ACL	I	Small	54	180	177.00	-3	0.98	197.53
ACL	E	Small	14	180	118.57	-61.43	0.66	86.59

Online Course Representatives

Online Course Representatives had four primary duties: facilitating the course group (45% of contracted time), coordinating with the course council (30% of contracted time), periodically evaluating the quality of grading and content-related feedback provided by instructors to students using course rubrics and helping with grade norming (20% of contracted time), and redirecting stakeholder requests to appropriate channels (5% of contracted time). A

breakdown of expected time and actual time in their four primary duties based off contract size is demonstrated below (see Table 7).

Table 7

Comparison of Contracted Time with Average Time for OCRs

OCR Principal Duties	% Of Contracted Time	Group Size	N	Expected Time	Average Time Spent	Time Diff.	% Time Diff.	SD
Facilitating the course group	45 %	Micro	237	27	24.06	-2.94	0.89	28.49
		Small	46	54	50.51	-3.49	0.94	62.02
Coordinating with the course council	30%	Micro	237	18	34.68	16.68	1.93	69.67
		Small	46	36	31.99	-4.01	0.89	57.72
Periodically evaluating the quality of grading...	20%	Micro	237	12	3.28	-8.72	0.27	8.54
		Small	46	24	2.63	-21.37	0.11	7.35
Directing stakeholder requests to appropriate channels	5%	Micro	237	3	4.72	1.72	1.57	14.87
		Small	46	6	1.85	-4.15	0.31	6.86
Other	0	Micro	237	0	7.22	7.22		32.81
		Small	46	0	8.15	8.15		38.14

Micro Group OCRs

Based on the percentages presented in the job descriptions and the organizational expectation that a one-hour contract equates 240 minutes of work per week, time spent on OCR duties of micro groups was calculated using the formula Percent of Contracted Time (PCT) x Contract Size (CS) = Expected Time (ET). As OCRs of Micro groups receive a .25-hour contract for their duties, they were expected to spend 60 minutes weekly in their responsibilities. A comparison of Online Course Representative's estimated time based off the different duties in

their job description with the reported average time for these activities revealed that micro group OCRs spent 3 minutes less or 11% less time facilitating the course group than contracted.

Coordinating with the course council took 17 minutes more time or 193% of contracted time for this duty. Time spent periodically evaluating the quality of grading and feedback in their course group was 9 minutes less or 27% of contracted time. Directing stakeholder requests took about 2 minutes more time or 157% of contracted time. Activities related to their duties but unaccounted for in their contracts took about 7 minutes of time each week (see Table 7). Standard deviations for all categories showed that micro group OCRs varied widely in their time spent throughout the semester.

Small Group OCRs

Based on the percentages presented above for OCRs' contracted time and the organizational expectation that a one-credit hour contract equates 240 minutes of work per week (see Table 7), OCR duties of small groups were calculated using the formula Percent of Contracted Time (PCT) x Contract Size (CS) = Expected Time (ET). As OCRs of small groups receive a .5-hour contract for their duties, they were expected to spend 120 minutes weekly in their responsibilities. A comparison of Online Course Representative's estimated time based off the different duties in their job description with the reported average time for these activities revealed that OCRs of Small groups spent 3 minutes less or 6% less time facilitating the course group. Coordinating with the course council took 4 minutes less or 89% of contracted time. Time spent periodically evaluating the quality of grading and feedback in their course group was 21 minutes less or 21% of contracted time. Directing stakeholder requests took about 4 minutes less time or 31% of contracted time. Activities related to their duties but unaccounted for in their contracts took about 8 minutes of time each week (see Table 7). Standard deviations for all

categories showed that small group OCRs varied widely in their time spent throughout the semester in most areas. Small group OCRs were most consistent in the amount of time spent directing stakeholder requests and periodically evaluating the quality of grading and feedback in their course group.

Assistant Course Leads

Assistant Course Leads had four primary duties: facilitating the course group (25% of contracted time), proactively improving course design & delivery (55% of contracted time), applying the Adaptive Course Fix Model (15% of contracted time), and directing stakeholder requests to the appropriate channels (5% of contracted time). In addition, their job descriptions indicate they must represent the course lead's vision for the course to the course group, but this is not calculated into their contracts. A breakdown of how much time is expected in their four primary duties plus representing the course lead based off contract size is demonstrated below (see Table 8).

Table 8

Comparison of Contracted Time with Average Time for ACLs

ACL Principal duties	% Of Contract	Group Size	N	Expected Time	Average Time	Time Diff.	% Time Diff.	SD
Facilitating the course group	25	Micro	7	30	43.25	13.25	1.44	69.70
		Small	3	45	144	99	3.2	10.39
Proactively improving course design & delivery	55	Micro	7	66	73.68	7.68	1.12	66.40
		Small	3	99	68.33	-30.67	-0.69	82.51
Applying the Adaptive Course Fix model	15	Micro	7	18	11.54	-6.46	-0.64	15.65
		Small	3	27	18.33	-8.67	-0.68	27.54

Table 8 continued

ACL Principal duties	% Of Contract	Group Size	N	Expected Time	Average Time	Time Diff.	% Time Diff.	SD
Directing stakeholder requests to appropriate channels	5	Micro	5	6	3	-3	-0.50	4.47
		Small	2	9	16.50	7.5	1.83	19.09
Representing the CL to the Course Group	0	Micro	7	0	18.61	18.61		25.91
		Small	3	0	35.67	35.67		28.75
Other Duties	0	Micro	7	0	3.71	3.71		9.39
		Small	3	0	0	0		0

Micro Group ACLs

Based on these percentages and the organizational expectation that a one-credit hour contract equates 240 minutes of work per week, ACL duties of micro groups were calculated using the formula Percent of Contracted Time (PCT) x Contract Size (CS) = Expected Time (ET). As ACLs of Micro groups receive a .5-hour contract for their duties, they were expected to spend 120 minutes weekly in their responsibilities. Facilitating the course group took about 13 minutes more or 44% more time than contracted, and they spent about 8 minutes more or 12% more time proactively improving course design and delivery than contracted. Making fixes to the course took 6 minutes less or 36% less time than contracted while directing stakeholder requests took about 3 minutes less or 50% less time than contracted. Additionally, duties not accounted for in ACL contracts like representing the Course Lead's vision to the course group and other duties took about 22 minutes total (see Table 8) beyond what was contracted. Standard deviations for all categories showed that micro group ACLs varied widely in their time spent throughout the semester.

Small Group ACLs

Based on the percentages presented above for ACLS' contracted time and the organizational expectation that a one-credit hour contract equated 240 minutes of work per week, ACL duties of small groups were calculated using the formula Percent of Contracted Time (PCT) x Contract Size (CS) = Expected Time (ET). As ACLs of small groups receive a .75-hour contract for their duties, they were expected to spend 180 minutes weekly in their responsibilities. A breakdown of how much time was expected in their four primary duties based off contract size is demonstrated in Table 8 (see Table 8). Facilitating the course group took about 99 minutes more or 320% of contracted time. They spent about 31 minutes less or 31% less time proactively improving course design and delivery than contracted. Making fixes to the course took about 6 minutes less or 36% less time than contracted, and directing stakeholder requests took about 8 minutes more time or 183% of contracted time. Additionally, Representing the Course Lead's vision to the course group added 36 minutes to their contracted time (see Table 8). Standard deviations for all categories showed that small group ACLs varied widely in their time spent throughout the semester.

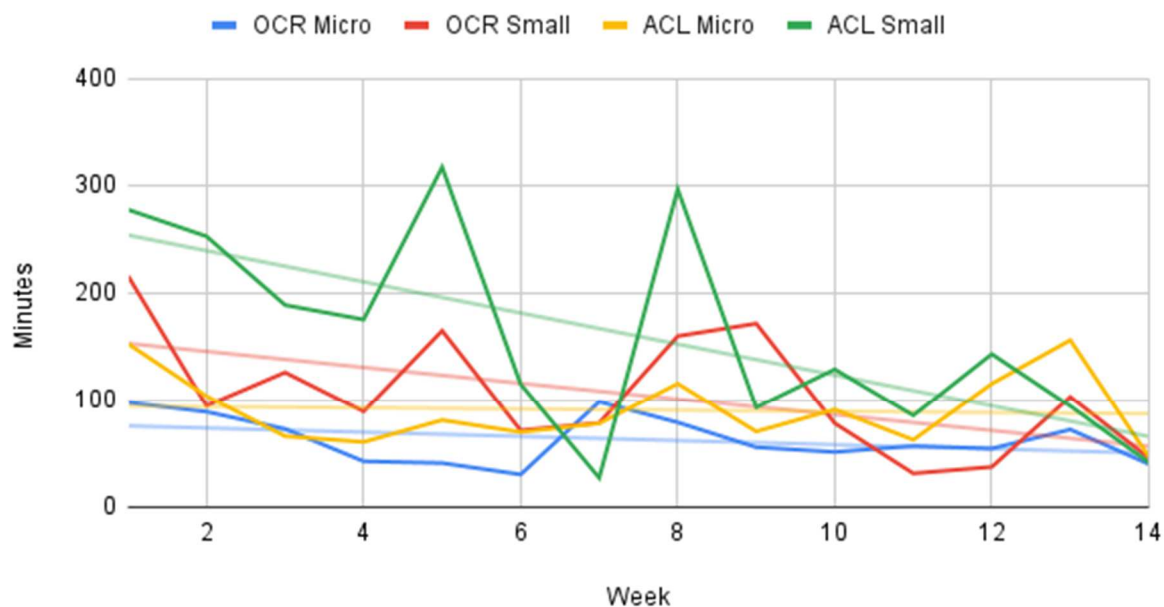
Week by Week Activity of Remote Leaders

Time reported on remote leaders' duties changed on a week-by-week basis. (see Table 9). Standard deviations for each group's weekly activity revealed vast inconsistency among remote leaders in the time they spent in any given week.

Table 9*Remote Leaders' Week by Week Average Time*

Week	OCR Micro			OCR Small			ACL Micro			ACL Small		
	Mean	N	SD	Mean	N	SD	Mean	N	SD	Mean	N	SD
1	97.90	20	59.24	215.83	6	271.80	152.93	7	158.02	277.33	3	151.27
2	89.09	11	74.56	95.00	2	49.50	104.00	6	135.43	252.60	5	218.38
3	73.02	14	56.59	126.25	4	75.65	66.24	10	50.52	188.83	6	160.04
4	42.76	17	27.49	89.50	2	41.72	60.75	8	45.83	175.38	8	65.00
5	41.11	18	38.75	165.00	5	76.08	81.32	14	104.14	316.67	6	422.58
6	30.47	17	36.84	71.67	3	42.52	70.08	10	79.08	115.29	7	77.14
7	98.54	12	165.96	78.57	2	101.02	78.18	11	68.35	27.33	3	21.94
8	79.14	25	127.20	160.00	4	197.61	115.78	9	177.01	296.25	4	261.64
9	55.86	21	45.48	171.67	3	133.82	70.63	8	58.26	93.40	5	85.98
10	51.42	12	99.98	78.33	3	79.74	91.25	8	113.35	129.13	6	130.18
11	57.00	17	58.68	31.45	3	46.42	62.82	11	35.72	85.75	4	67.12
12	54.69	24	63.07	37.50	2	3.54	115.71	7	100.18	143.33	3	137.96
13	72.41	16	112.47	103.33	3	114.27	156.25	7	296.43	94.60	5	63.86
14	40.21	13	79.82	45.84	4	46.15	47.00	7	39.30	41.67	3	10.41
Total	62.70	237	81.83	117.80	46	132.37	87.83	123	114.87	164.97	68	181.35

A graph of the different types of remote leaders' week by week activity with trendlines for each role visualizes how different types of remote leaders vary in their time over the semester (see Figure 3). ACLs of small groups had the steepest decline of activity from start of semester to end of semester, while ACLs of micro groups had the most consistent amount of activity over the entire semester. Both ACL and OCR small group leaders had the greatest fluctuation week by week, represented by the peaks and valleys of their lines below (see Figure 3).

Figure 3*Week by Week Comparison of Remote Leader***Week by Week Comparison of Time by Role****Week by Week Activity of Micro Group OCRs**

Time reported on Online Course Representatives duties outlined in the job descriptions for micro groups changed on a week-by-week basis. OCR time was front loaded in the semester with heightened activity mid-semester (see Table 10).

Table 10*Micro Group OCR Reported Time Week by Week*

Week		Facilitate	Evaluate	Coordinate	Direct	Other
1	Mean	63.30	6.50	29.89	5.01	21.30
	N	20.00	20.00	20.00	20.00	20.00
	SD	45.15	16.07	30.39	16.93	89.31
2	Mean	40.45	2.73	38.18	3.18	0.00
	N	11.00	11.00	11.00	11.00	11.00

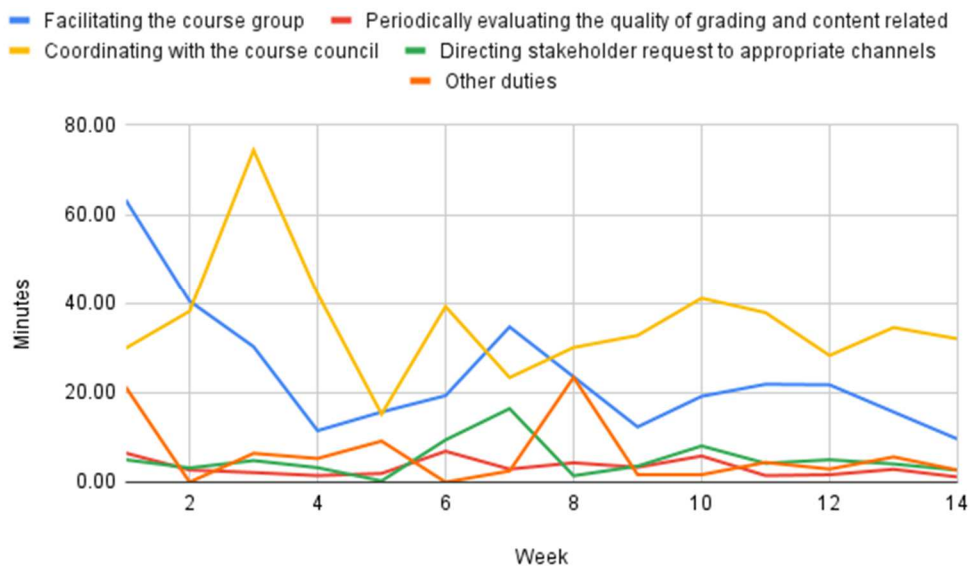
Table 10 continued

Week		Facilitate	Evaluate	Coordinate	Direct	Other
2	SD	35.25	6.47	37.37	10.55	0.00
3	Mean	30.25	2.14	74.39	4.82	6.43
	N	14.00	14.00	14.00	14.00	14.00
	SD	36.37	5.50	207.81	10.76	24.05
4	Mean	11.50	1.47	42.13	3.24	5.29
	N	17.00	17.00	17.00	17.00	17.00
	SD	8.22	3.86	103.06	12.11	15.86
5	Mean	15.67	1.94	15.20	0.28	9.17
	N	18.00	18.00	18.00	18.00	18.00
	SD	14.05	7.10	39.45	1.18	28.71
6	Mean	19.32	6.90	39.22	9.44	0.00
	N	17.00	17.00	17.00	17.00	17.00
	SD	20.79	11.78	43.07	20.52	0.00
7	Mean	34.67	2.92	23.35	16.42	2.50
	N	12.00	12.00	12.00	12.00	12.00
	SD	31.03	5.82	27.97	41.12	8.66
8	Mean	23.52	4.32	30.07	1.43	23.40
	N	25.00	25.00	25.00	25.00	25.00
	SD	21.62	8.10	32.28	3.38	44.41
	Mean	12.33	3.33	32.76	3.57	1.67
9	N	21.00	21.00	21.00	21.00	21.00
	SD	12.35	10.65	34.78	7.77	5.77
	Mean	19.17	5.83	41.08	8.08	1.67
10	N	12.00	12.00	12.00	12.00	12.00
	SD	29.88	11.04	57.64	18.05	5.77
11	Mean	21.87	1.47	37.85	4.21	4.41
	N	17.00	17.00	17.00	17.00	17.00
	SD	26.00	4.24	71.39	11.04	12.73
12	Mean	21.74	1.67	28.29	5.00	2.92
	N	24.00	24.00	24.00	24.00	24.00

Table 10 continued

Week		Facilitate	Evaluate	Coordinate	Direct	Other
12	SD	24.87	4.08	43.70	14.30	10.83
13	Mean	15.67	2.88	34.53	4.06	5.63
	N	16.00	16.00	16.00	16.00	16.00
	SD	15.54	8.26	38.11	7.79	22.50
14	Mean	9.62	1.15	32.00	2.69	2.69
	N	13.00	13.00	13.00	13.00	13.00
	SD	10.50	3.00	64.69	8.32	6.65

A graph of the different duties of OCRs leading micro groups week by week with trendlines for each duty visualizes how these OCRs were primarily focused on facilitating their course groups and coordinating with the course council. It also reveals that their efforts facilitating the course group were busiest in the start of the semester but quickly dropped as the semester progresses. Additionally, coordinating with the course council remained a relatively time intensive duty for these OCRs (see Figure 4).

Figure 4*OCR Micro Group Principal Duties Week by Week***Week by Week Activity of Small Group OCRs**

Time reported on Online Course Representatives duties for small groups changed on a week-by-week basis. OCR time was front loaded in the semester with heightened activity mid-semester (see Table 11).

Table 11*Small Group OCR Reported Time Week by Week*

Week		Facilitate	Evaluate	Coordinate	Direct	Other
1	Mean	107.50	6.67	31.67	0.00	0.00
	N	6.00	6.00	6.00	6.00	6.00
	SD	94.17	9.83	18.35	0.00	0.00
2	Mean	75.00	5.00	10.00	10.00	50.00
	N	2.00	2.00	2.00	2.00	2.00

Table 11 continued

Week		Facilitate	Evaluate	Coordinate	Direct	Other
2	SD	49.50	7.07	14.14	14.14	70.71
	Mean	50.00	10.00	52.50	3.75	0.00
3	N	4.00	4.00	4.00	4.00	4.00
	SD	34.64	20.00	35.94	7.50	0.00
4	Mean	25.00	5.25	2.50	0.00	0.00
	N	2.00	2.00	2.00	2.00	2.00
5	SD	7.07	6.72	3.54	0.00	0.00
	Mean	40.00	0.00	22.00	1.00	48.00
6	N	5.00	5.00	5.00	5.00	5.00
	SD	22.64	0.00	31.34	2.24	107.33
7	Mean	21.67	6.67	8.33	0.00	0.00
	N	3.00	3.00	3.00	3.00	3.00
8	SD	16.07	5.77	14.43	0.00	0.00
	Mean	9.27	0.00	5.00	0.00	0.00
9	N	2.00	2.00	2.00	2.00	2.00
	SD	1.03	0.00	7.07	0.00	0.00
10	Mean	66.25	0.00	16.25	0.00	1.25
	N	4.00	4.00	4.00	4.00	4.00
11	SD	92.86	0.00	11.09	0.00	2.50
	Mean	85.00	0.00	18.33	1.67	0.00
12	N	3.00	3.00	3.00	3.00	3.00
	SD	108.28	0.00	20.21	2.89	0.00
13	Mean	66.67	0.00	53.33	13.33	0.00
	N	3.00	3.00	3.00	3.00	3.00
14	SD	102.75	0.00	71.82	23.09	0.00
	Mean	10.27	0.00	121.67	0.00	0.00
15	N	3.00	3.00	3.00	3.00	3.00
	SD	9.84	0.00	206.42	0.00	0.00
16	Mean	20.00	0.00	17.50	0.00	0.00
	N	2.00	2.00	2.00	2.00	2.00

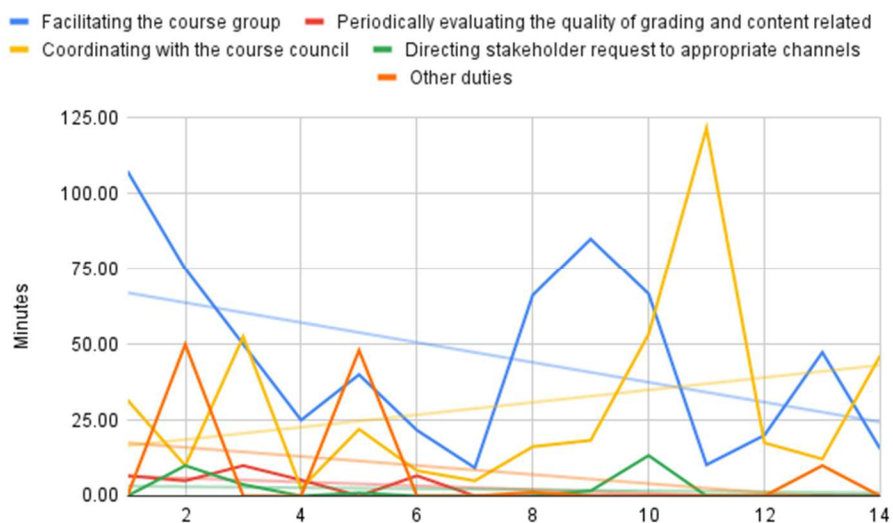
Table 11 continued

Week		Facilitate	Evaluate	Coordinate	Direct	Other
12	SD	14.14	0.00	10.61	0.00	0.00
13	Mean	47.33	0.17	12.18	0.00	10.00
	N	3.00	3.00	3.00	3.00	3.00
	SD	37.43	0.29	15.78	0.00	17.32
14	Mean	15.50	0.00	46.25	0.00	0.00
	N	4.00	4.00	4.00	4.00	4.00
	SD	18.65	0.00	40.70	0.00	0.00

A graph of the different duties of OCRs leading small groups week by week with trendlines for each duty visualizes how these OCRs were primarily focused on facilitating their course groups and coordinating with the course council. Additionally, coordinating with the course council in small course groups had a large peak in demand at the end of the semester, probably to adjust the course in preparation for the coming semester.

Figure 5

OCR Small Group Principal Duties Week by Week



Week by Week Activity of Micro Group ACLs

Time reported on Assistant Course Lead duties for micro groups changed on a week-by-week basis. ACL time was front loaded in the semester with heightened activity mid-semester and end of semester (see Table 12).

Table 12

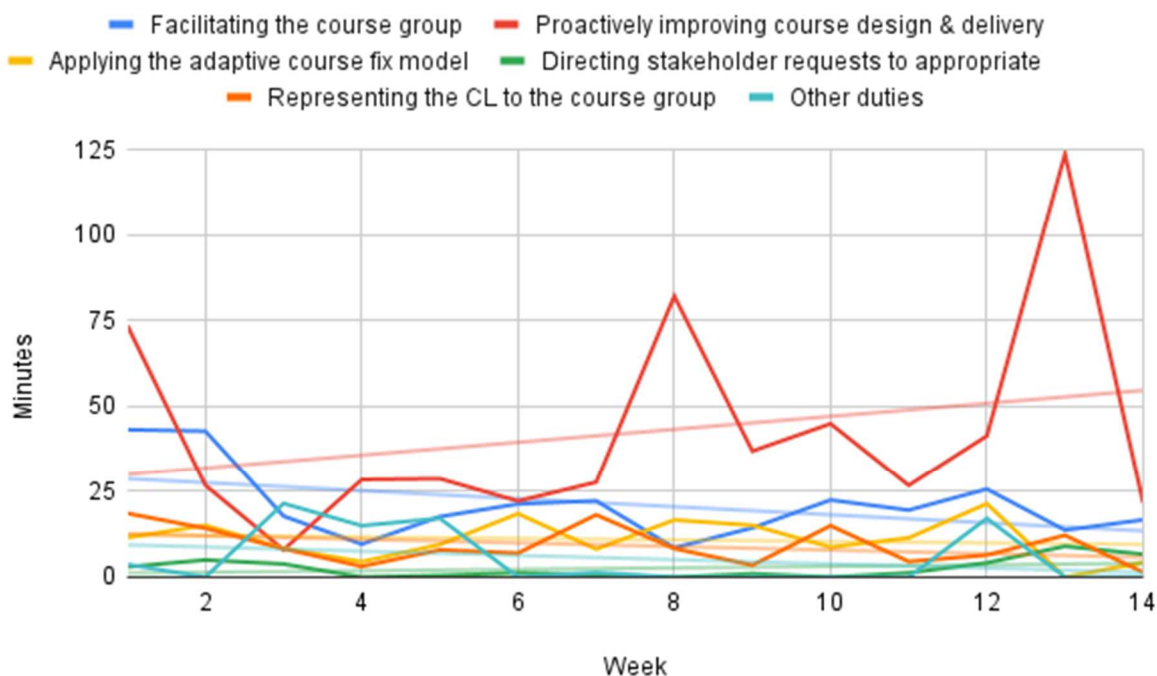
Micro Group ACL Reported Time Week by Week

Week		Facilitate	Improve	Fix	Direct	Represent	Other
1	Mean	43.25	73.68	11.54	3.00	18.61	3.71
	N	7.00	7.00	7.00	5.00	7.00	7.00
	SD	69.70	66.40	15.65	4.47	25.91	9.39
2	Mean	42.83	26.83	15.08	5.00	14.25	0.00
	N	6.00	6.00	6.00	6.00	6.00	6.00
	SD	40.79	41.30	25.04	8.37	30.02	0.00
3	Mean	17.80	7.97	8.40	3.83	8.28	21.50
	N	10.00	10.00	10.00	6.00	10.00	10.00
	SD	23.38	11.12	16.87	8.01	18.62	39.30
4	Mean	9.63	28.50	4.50	0.00	3.13	15.00
	N	8.00	8.00	8.00	4.00	8.00	8.00
	SD	5.58	35.93	8.40	0.00	3.72	42.43
5	Mean	17.64	28.75	9.50	0.50	7.93	17.14
	N	14.00	14.00	14.00	10.00	14.00	14.00
	SD	30.73	30.79	24.02	1.08	19.36	64.14
6	Mean	21.35	22.23	18.50	1.25	7.00	0.00
	N	10.00	10.00	10.00	8.00	10.00	10.00
	SD	24.54	32.76	20.01	3.54	13.98	0.00
7	Mean	22.27	27.73	8.18	0.71	18.18	1.36
	N	11.00	11.00	11.00	7.00	11.00	11.00
	SD	26.30	27.69	8.45	1.89	34.59	4.52

Table 12 continued

Week		Facilitate	Improve	Fix	Direct	Represent	Other
8	Mean	8.56	82.22	16.67	0.00	8.33	0.00
	N	9.00	9.00	9.00	7.00	9.00	9.00
	SD	11.29	156.13	26.46	0.00	11.73	0.00
9	Mean	14.38	37.00	15.13	1.00	3.50	0.00
	N	8.00	8.00	8.00	5.00	8.00	8.00
	SD	20.08	32.89	21.28	2.24	4.66	0.00
10	Mean	22.50	45.00	8.75	0.00	15.00	0.00
	N	8.00	8.00	8.00	5.00	8.00	8.00
	SD	40.00	48.84	11.57	0.00	25.50	0.00
11	Mean	19.55	26.82	11.45	1.25	4.55	0.00
	N	11.00	11.00	11.00	4.00	11.00	11.00
	SD	25.75	29.94	13.57	2.50	6.11	0.00
12	Mean	25.71	41.43	21.43	4.17	6.43	17.14
	N	7.00	7.00	7.00	6.00	7.00	7.00
	SD	30.06	43.85	26.73	4.92	11.07	37.29
13	Mean	13.71	123.89	0.00	9.00	12.21	0.00
	N	7.00	7.00	7.00	5.00	7.00	7.00
	SD	19.40	294.02	0.00	13.42	19.92	0.00
14	Mean	16.71	21.71	4.29	6.67	1.43	0.00
	N	7.00	7.00	7.00	3.00	7.00	7.00
	SD	20.79	25.99	11.34	5.77	2.44	0.00

A graph of the different duties of ACLs leading micro groups week by week with trendlines for each duty visualizes how these ACLs were primarily focused on making fixes to the master and blueprint courses and proactively improving course design and delivery. Like OCRs of micro groups, ACLs of Micro groups were busy facilitating their course groups early in the semester, but by week 4 this duty becomes less urgent for the remaining weeks (see Figure 6).

Figure 6*ACL Micro Group Principal Duties Week by Week***Week by Week Activity of Small Group ACLs**

Time reported on Assistant Course Lead duties for small groups changed on a week-by-week basis. ACL time was front loaded in the semester with heightened activity mid-semester (see Table 13).

Table 13*Small Group ACL Reported Time Week by Week*

Week		Facilitate	Improve	Fix	Direct	Represent	Other
1	Mean	144.00	68.33	18.33	16.50	35.67	0.00
	N	3.00	3.00	3.00	2.00	3.00	3.00
	SD	10.39	82.51	27.54	19.09	28.75	0.00

Table 13 continued

Week		Facilitate	Improve	Fix	Direct	Represent	Other
2	Mean	135.00	54.60	32.00	12.50	15.00	6.00
	N	5.00	5.00	5.00	4.00	5.00	5.00
	SD	113.74	69.61	66.11	15.00	18.71	13.42
3	Mean	54.67	24.33	46.67	4.80	24.17	35.00
	N	6.00	6.00	6.00	5.00	6.00	6.00
	SD	44.34	20.77	100.33	6.38	54.44	61.24
4	Mean	67.50	48.13	12.88	2.14	7.50	37.50
	N	8.00	8.00	8.00	7.00	8.00	8.00
	SD	36.65	44.15	13.54	2.67	11.34	71.26
5	Mean	158.33	108.33	6.67	7.00	37.50	0.00
	N	6.00	6.00	6.00	5.00	6.00	6.00
	SD	204.93	153.77	8.16	8.37	70.69	0.00
6	Mean	53.57	34.29	4.29	5.00	7.43	12.86
	N	7.00	7.00	7.00	4.00	7.00	7.00
	SD	49.30	41.58	11.34	5.77	13.34	34.02
7	Mean	14.00	6.67	3.33	10.00	0.00	0.00
	N	3.00	3.00	3.00	1.00	3.00	3.00
	SD	12.17	5.77	5.77	.	0.00	0.00
8	Mean	107.50	120.00	7.50	5.00	57.50	0.00
	N	4.00	4.00	4.00	3.00	4.00	4.00
	SD	137.39	63.77	9.57	8.66	60.21	0.00
9	Mean	45.40	19.00	9.00	6.67	16.00	0.00
	N	5.00	5.00	5.00	3.00	5.00	5.00
	SD	33.92	14.75	11.94	7.64	28.15	0.00
10	Mean	41.08	61.08	12.54	6.67	7.75	3.33
	N	6.00	6.00	6.00	3.00	6.00	6.00
	SD	46.46	111.79	24.01	5.77	11.55	8.17
11	Mean	32.50	30.75	17.50	2.50	3.75	0.00
	N	4.00	4.00	4.00	2.00	4.00	4.00

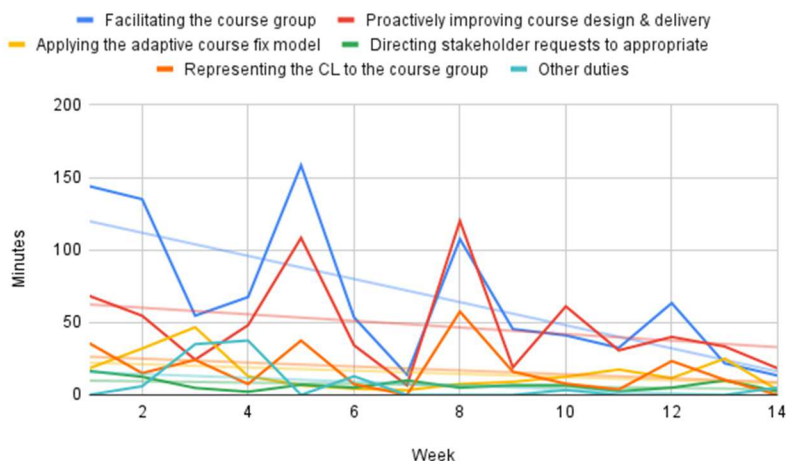
Table 13 continued

Week		Facilitate	Improve	Fix	Direct	Represent	Other
11	SD	17.08	41.74	15.00	3.54	4.79	0.00
12	Mean	63.33	40.00	11.67	5.00	23.33	0.00
	N	3.00	3.00	3.00	3.00	3.00	3.00
	SD	88.36	25.00	10.41	8.66	16.07	0.00
13	Mean	21.80	33.40	25.00	10.00	10.40	0.00
	N	5.00	5.00	5.00	2.00	5.00	5.00
	SD	19.65	19.31	24.24	7.07	12.38	0.00
14	Mean	13.33	18.33	3.33	2.50	0.00	5.00
	N	3.00	3.00	3.00	2.00	3.00	3.00
	SD	7.64	20.21	5.77	3.54	0.00	8.66

A graph of the different duties of ACLs leading small groups week by week with trendlines for each duty visualizes how these ACLs were primarily focused on facilitating the course group and proactively improving course design and delivery. For this group, these activities generally echoed each other throughout the semester (see Figure 7).

Figure 7

ACL Small Group Principal Duties Week by Week



Fall 2021 Course Group Survey

The Fall 2021 Course Group Survey is an organizationally generated survey used to collect data from course group members served by the remote leader and identify the activities course group members observed the remote leader performing in their duties. I averaged the number of reported instances of the different activities by remote leader type and group size in the time study and looked for alignment between the time spent and the number of instances of specific activities observed by group members. The alignment between observed activity and reported time confirmed the degree to which remote leader activity was noticed by those they serve and suggested the degree to which the role was being performed as it has been designed. In relation to RQ1, this confirmed the accuracy of the time study and further helped identify to what degree the new job duties had been institutionalized.

In the Fall 2021 survey, the most observed remote leader activities were facilitating a start of semester course group meeting (0.79), creating and participating in a course group chat (0.76), working with instructors on course improvements (0.68), making fixes in the course (0.67), and sharing feedback from instructors with the course council (0.58) (see Table 14). These observations aligned with OCR time reported performing duties of facilitating the course group and coordinating with the course council, which accounted for 75% of OCRs contracted

time and 79% of micro group OCRs and 86% of OCRs of small groups' average time, respectively (see Table 7). Likewise, facilitating the course group and proactively improving course design and deliver accounted for 80% of ACL contracted time and 76% of micro group ACL and 75% of small group ACLs' average time, respectively (see Table 8).

Similarly, the least observed remote leader activities were reviewing the quality of grading and feedback in the course (0.25), conducting grade norming activities (0.18), and following up with instructors out of sync with group performance (0.10) (see Table 14). These observations aligned with OCR time reported performing duties of periodically evaluating the quality of grading and content-related feedback provided by instructors to students using course rubrics and helping with grade norming, which accounted for 20% of OCRs contracted time and 11% and 5% of average time for micro and small group OCRs, respectively (see Table 7). Likewise, these activities were a subset of ACLs' duty to proactively improving course design & delivery, which accounted for 55% of ACL contracted time and 48% of their micro group ACL and 24% of small group average time (see Table 8). While not every instance of remote leader activity could be observed, that the most reported activities of ACLs were those most observed by instructors and those least reported by remote leaders were least observed by instructors suggested consistency among surveys and the reliability of the time study.

Table 14

Fall 2021 Course Group Survey Results

Remote Leader Activity	Spring 2021 % Instructor Observation (N=1260)	Fall 2021 % Instructor Observation (N=1140)	Difference between Fall and Spring Observations
------------------------	--	--	--

Facilitating a start of semester course group meeting	0.81	0.79	-0.03
Creating and participating in a course group chat	0.75	0.76	0.01
Working with instructors to identify and plan course improvements	0.68	0.68	0.00
Making course fixes when necessary	0.61	0.67	0.06

Table 15 continued

Remote Leader Activity	Spring 2021 % Instructor Observation (N=1260)	Fall 2021 % Instructor Observation (N=1140)	Difference between Fall and Spring Observations
Sharing feedback from instructors with the course council	0.58	0.58	0.00
Working with the course council to implement course improvement projects	0.53	0.55	0.02
Reviewing the quality of grading and feedback in the course	0.26	0.25	-0.01
Conducting grade norming activities	0.19	0.18	-0.01
Following up with instructors who are out of sync with group performance	0.14	0.10	-0.03
None of the above	0.06	0.05	-0.01

Comparison of data from Spring 2021 & Fall 2021 Course Group Surveys revealed course group satisfaction across semesters after the changes have been implemented. The difference between Spring 2021 & Fall 2021 data revealed minimal net changes on item level and total satisfaction. Instructor observation of remote leader activity in Spring and Fall 2021 semesters was nearly identical, with the greatest difference in observation being a 6% increase in instructors seeing their OCR or ACL making course fixes when necessary (see Table 14). This

lack of change from one semester to another suggested remote leaders' activity is relatively consistent across semesters since the organizational change had been implemented and the degree to which these roles have been institutionalized.

Fall 2021 Course Council Survey

The Course Council Survey was an organizationally generated survey that collected course lead observation of remote leaders' work. The survey was sent to course leaders in week 12 of the semester. Faculty who serve as course lead for multiple courses received one survey for each course. The survey received a 28% response rate (N=80) among course leads working with OCRs and a 42% response rate among course leads working with ACLs (N=58). Calculating the difference between the percent of OCRs and ACLs observed ($\text{OCR \% Observed} - \text{ACL \% Observed} = \text{\% Difference}$) revealed that ACLs were observed performing all activities more than OCRs. The greatest difference in observations (29%) was in making fixes to the blueprint course, while the most similar activity (2%) was in coordinating with the course lead (see Table 15).

Table 16

Comparison of Percent of Course Lead Observations of Remote Leader Activity

CL Observations of Remote Leader Activity	OCR % Observed (N=80)	ACL % Observed (N=58)	% Difference
Making fixes to the blueprint course	0.08	0.36	-0.29
Making fixes to the master course	0.35	0.57	-0.22
Organizing or conducting course group meetings	0.10	0.31	-0.21
Collaborating with online instructors to identify and plan course improvements	0.24	0.41	-0.18
Helping instructors find answers to their questions	0.28	0.43	-0.16
Sending information or announcements to your course group	0.29	0.43	-0.14
Working on or reviewing course improvements	0.41	0.55	-0.14
Reviewing course group performance data to identify and address instructors out of sync with the group	0.05	0.17	-0.12
Mentoring instructors regarding course-specific issues	0.23	0.34	-0.12

Evaluating the quality of grading and content-related feedback provided by instructors	0.11	0.21	-0.09
Meeting with or communicating with the course council regarding course improvements	0.56	0.66	-0.09
Conducting grade norming or other facilitation improvement activities	0.09	0.16	-0.07
Rerouting instructors' questions to the appropriate channels	0.18	0.24	-0.07
Responding to discussions in the course group chat	0.15	0.21	-0.06
Representing the course lead in communication or meetings to the course group	0.30	0.34	-0.04
Coordinating with the course lead via meeting, email, or chat	0.60	0.62	-0.02

Additionally, this data aligned with time study data that revealed ACLs were spending more time in their duties than their OCR counterparts in the same size course group (see Table 16). This alignment of course leads observing ACLs engaged in more activities than OCRs with ACLs reporting spending more time in their duties than OCRs of the same size groups suggested that the time study results were reliable because ACLs were contracted for more time than their OCR counterparts in course groups of the same size. In relation to RQ1, this confirmed the accuracy of the time study report and further helped identify to what degree the new job duties had been institutionalized.

Table 17
Comparison of Remote Leader Time by Role & Size of Group

<u>Role & Size of Group</u>	<u>Expected Weekly Time</u>	<u>Average Weekly Time</u>
OCR Micro	60	66
ACL Micro	120	107
OCR Small	120	145
ACL Small	180	167

Conclusion of Research Question 1

Research Question 1 was *how can confirmative evaluation be used to identify the degree to which organizational changes have been institutionalized?* The organizational change in this instance was the migration to course groups as the organizing structure for online instructors. The institutional concern was that remote leaders needed to be performing their new roles as defined by their job descriptions. Based on the findings of the time study, corroborated with the results of the course group and course council surveys, OCRs and ACLs were generally close to or over the time expected for their role and contract size. Assistant Course Leads of micro groups were least aligned with their job descriptions, performing 73% of the job time. ACLs of small groups were performing 92% of their job time. Online Course Representatives of micro groups were performing 105% of their job time while OCRs of small groups were performing 98% of their job time.

While total time on duty was close to contracted time for all OCRs and ACLs of small groups, ACLs of micro groups were spending far less time than contracted. Additionally, time spent on different duties varied from what was expected by the organization. Online Course Representatives of micro groups spent far more time (1.93) coordinating with their course councils and far less time (0.27) periodically evaluating the quality of grading and content-related feedback provided by instructors to students using course rubrics and helping with grade norming. Online Course Representatives of small groups were fairly close to the time estimated for facilitating the course group (0.94) and coordinating with the course council (0.89) while largely ignoring their duty to periodically evaluate the quality of grading and content-related feedback provided by instructors to students using course rubrics and helping with grade norming (0.11).

ACLs of micro groups spent more time facilitating the course group (1.44) and fairly close to the contracted time for proactively improving course design and delivery and facilitating the course group (1.12), but they spent far less time applying the adaptive fix model (0.64). Assistant Course Leads of small groups were spending far less time proactively improving course design and delivery (0.69) and applying the adaptive course fix model (0.69), but they were spending far more time facilitating the course group (3.20) and directing stakeholder requests to the appropriate channels (1.83).

Further, week by week analysis of the different roles by type and size of group showed that all roles found the first two weeks of the semester busiest with a surge of activity mid semester and another surge at the end of the semester. Assistant Course Leads of small groups had the greatest difference between start of semester and end of semester busyness. OCRs of small groups and ACLs of micro groups were relatively consistent with each other over the semester, and given they are contracted for the same amount of time, this consistency was encouraging. Online Course Representatives of micro groups were most consistent in their week-by-week activity.

Based on this analysis, it appeared that remote leaders needed continued work on aligning their efforts with their job descriptions. Online Course Representatives were spending more time than contracted and ACLs were spending less time than contracted. ACLs of small groups were spending far too much time facilitating their course groups. To align with their job descriptions, OCRs needed to increase efforts in periodically evaluating the quality of grading and content-related feedback provided by instructors to students using course rubrics and helping with grade norming. ACLs needed to focus more on proactively improving course design and delivery. Both OCRs and ACLs could have spent less time with their course groups. Further, standard

deviations in nearly every analysis revealed vast inconsistency within types of remote leaders, suggesting that the activity of remote leaders by group varied widely. As an example of confirmative evaluation, this time study has revealed where total time and activity on specific duties for remote leaders needed to be addressed. At phase one of this case study, it is apparent that there was a gap between the job assigned and what was being done by remote leaders, but it had not yet revealed the cause of the gap or the degree to which this gap matters for the organization, which is addressed in Research Question 3.

Research Question 2

Research Question 2 was *to what degree can confirmative evaluation identify the effects of changes in continuous improvement efforts?* To answer this question, I searched for improved course council and course group satisfaction with remote leaders' performance. A positive difference in data from course group surveys or course council surveys suggested growing satisfaction and subsequent value from the new model, negative difference suggested decreasing satisfaction and subsequent decreased value from the new model, and neutral difference suggested no change in satisfaction or subsequent value as it applied to course group or course council satisfaction.

Course Group Survey

Four questions on the course group survey were relevant to instructor satisfaction with their course group and remote leaders. Questions were asked on 5-point scales ranging from *Not at all well* to *Extremely well*. Instructors indicated overall satisfaction in all four categories with minimal difference between semesters (see Table 17), suggesting no subsequent change for course groups in satisfaction with remote leader activity and participation in the new organizational model after it was implemented.

Table 18*Instructor Satisfaction with Course Groups*

Question	Fall 2021 Course Group Survey Very/Extremely Well Responses (N=1140)	Spring 2021 Course Group Survey Very/Extremely Well Responses (N=1260)	Difference
How comfortable do you feel about bringing course-specific questions to your course group?	0.847	0.862	0.015
When you bring questions to your course group, how well does the group provide useful answers?	0.796	0.803	0.007
How well did the start of semester course group meeting make use of your time?	0.650	0.656	0.006
How well does the course group chat facilitate course improvement discussion?	0.638	0.636	-0.002

Course Council Survey

The Course Council Survey asked Likert-type satisfaction questions of course leads regarding their satisfaction with their Assistant Course Leads and observations of how ACLs changed the dynamics of how the course council and course group functions. A comparison of Winter 2021 survey results (when the model was first implemented) with Fall 2021 results showed the subsequent change in satisfaction and observed increased value of ACLs since the model's implementation in Winter 2021. A comparison of the median response selected suggested no change in course lead overall perception of ACLs impact, but a closer examination of question responses revealed noteworthy changes (see Table 18). For example, for the question *How has having an ACL impacted collaboration and cooperation among Course Council members?* 36% of course leads reported collaboration and cooperation among course council

members was much improved in Winter 2021 while 70% of course leads reported the same in Fall 2021. Course leads reporting this topic was *somewhat improved* or *unaffected* significantly decreased between semesters while those reporting *somewhat worse* increased from zero to 12%. It appears that there was a polarizing shift among course leads' opinions regarding ACLs' impact in this area, with many more course leads viewing ACLs favorably in this area. Similarly, course leads demonstrated modestly shifting comfort with ACLs making changes to the course master, shifting patterns of communication frequency among the course council and feedback from online instructors, and overall course lead satisfaction with their role on the course council. Conversely, there was no perceived change from course leads in course maintenance or their desire to be engaged with the online instructors (see Table 18).

Table 19

Comparison of Course Lead Perception of ACL Value Winter 2021 and Fall 2021

Course Council Survey Question	Winter 2021 (N=46)	Fall 2021 (N=151)	Difference
How has having an ACL impacted collaboration and cooperation among Course Council members?	Median		
	4.0	4.5	0.5
Much improved	0.36	0.70	0.33
Somewhat improved	0.27	0.00	-0.27
Unaffected	0.36	0.18	-0.18
Somewhat worse	0.00	0.12	0.12
Much worse	0.00	0.00	0.00
How comfortable are you with the ACL making changes directly to the course master?	Median		
	4.00	4.00	0.0
Extremely	0.26	0.31	0.05
Very	0.44	0.33	-0.11
Moderately	0.13	0.22	0.10
Slightly	0.10	0.10	0.00
Not at all	0.08	0.04	-0.04

Table 18 continued

Course Council Survey Question	Winter 2021 (N=46)	Fall 2021 (N=151)	Difference
How frequently have you met with or communicated with the Assistant Course Lead to discuss changes to the course?	Median		
	2.00	2.00	0.0
Many times a day	0.00	0.00	0.00
Daily	0.00	0.00	0.00
A couple time a week	0.00	0.00	0.00
Weekly	0.31	0.16	-0.14
Less than once a week	0.50	0.72	0.22
Never	0.19	0.12	-0.08
How frequently have you met with or communicated with the entire Course Council to discuss changes to the course?	Median		
	2.00	2.00	0.0
Many times a day	0.00	0.00	0.00
Daily	0.00	0.00	0.00
A couple time a week	0.00	0.00	0.00
Weekly	0.09	0.09	-0.01
Less than once a week	0.91	0.81	-0.10
Never	0.00	0.11	0.11
How has having an ACL impacted the maintenance of the course?	Median		
	4.0	4.0	0.0
Much Better	0.45	0.46	0.00
Somewhat better	0.36	0.33	-0.04
Unaffected	0.18	0.22	0.04
Somewhat worse	0.00	0.00	0.00
Much worse	0.00	0.00	0.00
How has having an ACL impacted the feedback you receive from online instructors reporting needed fixes?	Median		
	3.0	3.0	0.0
Decrease Greatly	0.11	0.07	-0.04
Decrease somewhat	0.08	0.15	0.07
No change	0.61	0.39	-0.21
Increase somewhat	0.21	0.30	0.09
Increase greatly	0.00	0.09	0.09

Table 18 continued

Course Council Survey Question	Winter 2021 (N=46)	Fall 2021 (N=151)	Difference
How has having an ACL impacted the feedback you receive from online instructors suggesting course improvements?	Median 3.0	3.0	0.0
Decrease Greatly	0.08	0.02	-0.06
Decrease somewhat	0.05	0.11	0.06
No change	0.58	0.40	-0.18
Increase somewhat	0.24	0.36	0.12
Increase greatly	0.05	0.11	0.06
How has having an ACL impacted the feedback you receive from online instructors relating positive experiences with the course?	Median 3.0	3.0	0.0
Decrease Greatly	0.03	0.00	-0.03
Decrease somewhat	0.03	0.07	0.04
No change	0.68	0.52	-0.16
Increase somewhat	0.24	0.27	0.04
Increase greatly	0.03	0.14	0.11
How has having an ACL impacted the feedback you receive from online instructors relating negative experiences with the course?	Median 3.0	3.0	0.0
Decrease Greatly	0.05	0.00	-0.05
Decrease somewhat	0.03	0.14	0.11
No change	0.74	0.64	-0.10
Increase somewhat	0.18	0.14	-0.05
Increase greatly	0.00	0.09	0.09
How has having an ACL impacted the feedback you receive from online instructors on other topics?	Median 3.0	3.0	0.0
Decrease Greatly	0.00	0.00	0.00
Decrease somewhat	0.04	0.11	0.07
No change	0.92	0.79	-0.14
Increase somewhat	0.04	0.07	0.03
Increase greatly	0.00	0.04	0.04

Table 18 continued

Course Council Survey Question	Winter 2021 (N=46)	Fall 2021 (N=151)	Difference
How has having an ACL impacted your direct interactions with online instructors?	Median 3.0	3.0	0.0
A lot more	0.22	0.21	-0.01
Slightly more	0.22	0.36	0.13
Adequate	0.00	0.00	0.00
Slightly less	0.33	0.29	-0.05
A lot less	0.22	0.14	-0.08
Would you like greater opportunity to directly interact with online instructors?	Median 3.0	3.0	0.0
A lot more	0.00	0.00	0.00
Slightly more	0.17	0.18	0.01
Adequate	0.65	0.60	-0.05
Slightly less	0.00	0.04	0.04
A lot less	0.17	0.18	0.01
How satisfied are you with your role on the Course Council, with the inclusion of an ACL this semester?	Median 4.0	4.0	0.0
Extremely	0.15	0.33	0.17
Very	0.59	0.41	-0.18
Moderately	0.10	0.17	0.07
Slightly	0.08	0.07	-0.01
Not at all	0.08	0.02	-0.06

Conclusion of Research Question 2

Research Question 2 was *to what degree can confirmative evaluation identify the effects of changes in continuous improvement efforts?* The analysis of online instructors' satisfaction with the new organizational model between Spring 2021 and Fall 2021 showed no significant change in satisfaction, suggesting the model had remained relatively stable for instructors since its inception. Additionally, instructors indicated they were generally empowered to ask questions

and confident that they would receive useful feedback from their colleagues in their course groups, suggesting the course groups are adequately meeting instructors' support needs. Start of semester course group meetings had been overall acceptable, but their quality and usefulness could be further evaluated. Likewise, course group chats appeared to be overall useful to instructors, but they received the lowest evaluations from online instructors.

A comparison of Winter 2021 and Fall 2021 Course Council Survey data revealed course leads identifying both growing value in ACLs (collaboration and cooperation, increased feedback from online instructors, overall satisfaction in course leads' role) and increasing concern about ACLs (making fixes to master courses, less communication among the course council) while unchanged or ambivalent perception in other areas (impact on course maintenance, course lead direct interaction with online instructors).

As confirmative evaluation is about determining the degree to which implemented changes have continued to be effective and these changes' overall value, in this case confirmative evaluation has identified effects of changes in the online instructor organizational model and shows both value added and unrealized potential that can be sought through future interventions. Like Research Question 1, Research Question 2 has revealed gaps in the implementation of organizational changes at this institution, but it has not yet revealed the cause of the gap or the degree to which this gap matters for the organization, which will be addressed in Research Question 3.

Research Question 3

Research Question 3 is *how does data from confirmative evaluation support continuous improvement efforts to reinforce organizational changes?* To address this research question, I used the results from Research Question 1 and Research Question 2 to identify gaps in remote leader performance by identifying the average time spent by remote leaders based on the type of group and group size. I then determined the degree to which course councils and course groups were satisfied with remote leader performance and their experiences in this new model by comparing results from Research Question 1 and Research Question 2 to identify discrepancies between course group and council overall satisfaction and remote leaders' reported time (see Table 19). Incongruence between the time study and course group observations suggested a disconnect between remote leaders and those they served in the course group, identifying another gap in the adoption of the new leadership model. In total, 16 gaps were quantified in this round of confirmative evaluation. I then hypothesized on what might be causing these gaps based on the data collected and my knowledge of how the organization operates and identified potential areas of investigation for a gap analysis that would be the next step in the continuous improvement process.

Table 20*Gaps in Remote Leader Performance and Areas of Investigation*

Remote Leader	Observed Gaps	Areas of Investigation
All Groups	Standard deviations revealed overall and weekly inconsistency within each remote leader type	Remote leaders are not consistently performing the work outlined for them, Course-specific needs vary widely, Job descriptions do not match the specific roles
Institution E Micro Group ACLs	12% less time in total duties than Institution I Micro Group ACLs	Job description is not aligned to duties at Institution E
Institution E Small Group ACLs	32% less time in total duties than Institution I Small Group ACLs	Job description is not aligned to duties at Institution E
Micro Group OCRs	193% of contracted time coordinating with course councils	Chronic course issues, course development process
	27% of contracted time for evaluating grading and feedback quality	Lack understanding of how to do this, missing required systems permissions, relatively new duty for OCRs, no follow up
Small Group OCRs	11% of contracted time for evaluating grading and feedback quality	Lack understanding of how to do this, missing required systems permissions, relatively new duty for OCRs, no follow up
Micro Group ACLs	27% less total time than contracted	Insufficient work compared to contract
	144% of contracted time for facilitating the course group	Job descriptions needs further refinement, ACLs need to focus on other less obvious duties
Micro Group ACLs	64% of contracted time making fixes	Insufficient work compared to contract

Table 19 continued

Remote Leader	Observed Gaps	Areas of Investigation
Small Group ACLs	320% of contracted time for facilitating the course group	Job descriptions needs further refinement, most comfortable role for ACLs, course groups have more questions than they should, new instructors take extra time
	69% of contracted time for improving course design & delivery	Job descriptions needs further refinement, lack understanding of how to do this, missing required systems permissions, uncomfortable role for ACLs, no follow up
	68% of contracted time for making fixes to the course	Job descriptions needs further refinement, lack of confidence making fixes, no follow up
All Course Groups	35% of online instructors do not find the start of semester course group meeting beneficial	Lack of preparation by remote leader, misaligned meeting with actual instructor needs, too few instructors participating in the meeting
All Course Groups	36% of online instructors do not find the course group chat beneficial	Collaboration is unnecessary for some instructors, too few instructors participating in the chat, chat is filled with non-course related discussion
All ACLs	Course leads have increasing hesitance about ACLs making fixes to master courses	Course Leads lack of understanding about what ACLs can fix without their permission, inadequate communication among the course council
All ACLs	There is less communication among the course council than before	ACLs are not collaborating adequately with their course leads, changing required frequency of course council meeting

In terms of confirmative evaluation's role in continuous improvement processes, this case has demonstrated that it fills a gap between summative and formative assessment by identifying the reality at the institution compared to the idealized version of the new organizational model.

This confirmative evaluation study suggested 16 gaps between what is expected of the new online instructor organizational model and reality. It also identified areas worth investigating for the gap analysis that would come from this evaluation positions the organization at the start of a new round of continuous improvement to prioritize and address these gaps. The data from this confirmative evaluation could be directly used as the organization moves into its next efforts to improve how the organization functions.

Conclusion of Findings

This instrumental case study was an application of confirmative evaluation in a department of online learning that serves three higher education institutions in a private educational system. It sought to help illuminate our understanding of how confirmative evaluation works in reality and was guided by three questions:

1. How can confirmative evaluation be used to determine the degree to which organizational changes have been institutionalized?
2. To what degree can confirmative evaluation determine the effects of changes in continuous improvement efforts?
3. How does data from confirmative evaluation support continuous improvement efforts to reinforce organizational changes?

In this case, confirmative evaluation was conducted using a time study of remote leaders and extant data from the organization. The time study in this case determined the degree to which remote leaders' redefined work as outlined in their job descriptions has been institutionalized (RQ1) and found gaps in their performance of their roles. It also used course council and course group surveys to corroborate the results of this time study, adding to its reliability. Additionally, confirmative evaluation was implemented to determine the effects of changes (RQ2) in remote

leader work and found that since implementation of the new organizational model, course group satisfaction and course lead perception have remained consistent. It also found that the value course leads find with ACLs both increased and diminished, depending on the question asked. Finally, an analysis of the data (RQ3) produced in the case to answer research questions 1 and 2 produced the beginning of a gap analysis that the organization could then use to address issues in its implementation of the organizational change to further drive remote leaders towards the roles outlined for them in their job descriptions and to increase the value of these roles for course groups and course leads.

From the data presented in this case, it is apparent that confirmative evaluation can play an instrumental role in identifying the degree to which organizational changes have been institutionalized, identifying the effects of changes in continuous improvement efforts, and supporting the next steps in continuous improvement efforts. Change management often requires consistent, concentrated effort over time, and confirmative evaluation in this example has shown where the institution could continue to focus its energies if it desired to bring the reality of its new organizational model in line with its initial vision for organizing instructors. In the next chapter, I will apply the lessons learned in this instrumental case study to the performance improvement field's understanding of confirmative evaluation.

Limitations & Future Research

This instrumental case study, while showing one way confirmative evaluation can be conducted to fit organizational need and circumstance, brings with it important limitations. First, designed into the study was that sections of the entire remote leader structure were intentionally ignored. The time study excluded medium (21-40 instructors), large (41-60 instructors), and mega (+61 instructors) course groups because there were not enough course groups in these

categories to study remote leader work while sufficiently protecting participant anonymity. Further, these groups were not excluded from the extant course council or course group surveys prior to the fall 2021 semester, so they were not excluded from the Fall 2021 extant data collection. These remote leaders represented 3% of all remote leaders during the Fall 2021 semester, so they may have had a potentially small impact on these surveys' results. Different methods that would allow them to participate in future time studies would help address this.

This case's use of self-reported activity as the basis for the time study was another limitation. Ideally, an observer would track the work performed by the subjects of the time study (Overby, 1983), but this was impossible since remote leaders reside across the United States. While self-reporting was not the ideal approach, there was no better means to collect the data. In this case, weekly response rates revealed that not all remote leaders participated in the study, and there was not an undue pressure or incentive to do so, so it is likely a safe assumption that those who did participate did so in good faith and reported as accurately as they could.

Additionally, weekly responses to the time study could have been more robust. While the time study averaged 35 responses weekly, that was across all four types of remote leaders. Weekly response rates by group were generally at acceptable levels, but the n for each group was less than optimal (see Table 20). Likewise, the comparison of remote leaders between Institution I and Institution E relied on very small response rates for Institution E. Future research on remote leaders could find ways to improve reporting statistics or rework the matrix sampling to have remote leaders report more frequently to increase the response rate.

Table 21*Group N and Average Weekly Response Rates*

Type of Remote Leader	Average Weekly N	Average Weekly Response Rate
Micro Group OCR	17	34%
Small Group OCR	3	24%
Micro Group ACL	9	51%
Small Group ACL	5	55%

Another limitation in this study was that extant data of course council and course group surveys only examined satisfaction of these groups since the organizational change to the course group model had been implemented. Prior to this change there was no comparable data sets for either group and had there been this case could have made better comparisons of before, immediately after, and well after the organizational change to provide a truer representation of the full value of the organizational change and also highlighted the value of confirmative evaluation in examining a change well after it had been implemented. This highlights a challenge with conducting confirmative evaluation when it is not planned into a continuous improvement effort. Future cases that include planning for confirmative evaluation while changes are being implemented could be studied that would provide a fuller scope of evaluation.

Additionally, this case was intentionally limited to focus on one aspect of the organizational change: how remote leaders spent their time. This was one element of a larger organizational change that in addition to redefining remote leaders' work also modified full-time employee primary functions, evaluation procedures, HR functions, training, and onboarding. This case intentionally narrowed the scope of research for clarity for both the researcher and the study, but in doing so it ignored elements of the larger system within which the remote leaders

function. Incorporating that larger system may provide opportunity for a richer evaluation, but it also brings the potential of making the evaluation overly complex. Future instances of confirmative evaluation could take a broader scope to use confirmative evaluation more holistically than presented here if it can address the overall complexity of the task.

Finally, this case also intentionally ended at the beginning of the next round of a continuous improvement process by identifying gaps and areas of investigation for the gap analysis. In continuous improvement models like ISPIS's HPT model (see Figure 1), this case has shown where confirmative evaluation occurred during the implementation and maintenance phase and pulls the project back to the Performance Analysis or Need or Opportunity phase. Following this project or other projects into subsequent improvement projects and back to confirmative evaluation of those projects would demonstrate the differences among formative, summative, and confirmative evaluation to draw clearer distinctions among these types of evaluation for other practitioners while also showing the value of confirmative evaluation in a continuous improvement process as it sets the stage for the next iteration of interventions.

Even with these limitations, this case helped fill an important gap in our research around confirmative evaluation. It helped to define what confirmative evaluation really is and demonstrated some of the challenges of conducting confirmative evaluation faced by performance improvement practitioners. The next chapter will delve into a deeper discussion of what this instrumental case study revealed about confirmative evaluation and its implications for the field.

CHAPTER V

Discussion & Implications

In this case, the organizational changes at an online learning department serving three institutions of higher education were studied to identify how confirmative evaluation can be used in a continuous improvement process. The research questions guiding this instrumental case study were

1. How can confirmative evaluation be used to determine the degree to which organizational changes have been institutionalized?
2. To what degree can confirmative evaluation determine the effects of changes in continuous improvement efforts?
3. How does data from confirmative evaluation support continuous improvement efforts to reinforce organizational changes?

In response to RQ1, the time study of remote leaders' work demonstrated the degree to which they were performing their duties as outlined by their new job descriptions and identified quantifiable gaps in remote leader performance (see Table 19). Specifically, Online Course Representatives were spending more time than contracted while Assistant Course Leads were spending less time than contracted, and both roles could better align their time with specific duties in their job descriptions. Additionally, the review of course group surveys revealed consistency from one semester to another, suggesting the changes around remote leaders' work have been institutionalized in the organization. While the organizational change impacted more than just remote leaders and their roles, this study intentionally focused on these remote leaders.

In response to RQ2, the effects of changes with remote leader job duties were determined by observing the satisfaction over time of those whom remote leaders served. Online instructors'

satisfaction was measured using the organization's Course Group Survey over two semesters, and it showed nearly identical results from Spring 2021 to Fall 2021 (see Table 18). The results of these surveys demonstrated that online instructors were adequately receiving support within their course groups for course-related issues. Additionally, it demonstrated that course group start of semester meetings and group chats were less valuable to online instructors and could be an area of future improvement.

The effects of organizational changes on the course council were determined by observing the changes in the results of the organization's Course Council Survey from Winter 2021 to Fall 2021 semester. These surveys demonstrated increased satisfaction of course leads who have ACLs in this new model between semesters, especially in their satisfaction in their own roles and in the collaboration that occurs in the course council. It also revealed, interestingly, trends of concern regarding perceived decreasing communication among the course council generally and concern over ACLs making fixes in the course, suggesting further work is needed in these areas.

In response to RQ3, the review of how remote leaders spend their time and the analysis of course group and course council surveys revealed several gaps between the organization's idealized view of how remote leaders and course groups should function and their reality (see Table 20). The results of this confirmative evaluation were 16 quantified gaps in remote leader performance, along with potential areas for investigation for these gaps. Because this study was intentionally bounded to end with the identification of gaps and potential areas of investigation, it did not move into a deeper root cause analysis. It did, however, provide the organization a clear path forward for prioritizing and investigating these gaps in the next round of organizational improvement.

Instrumental case studies use cases to observe larger phenomenon presented in the case (Harling, 2002) and to provide insight or refine theory about that phenomenon (Baskarada, 2014). This instrumental case study used the case to explore confirmative evaluation. This chapter will unpack the discoveries about confirmative evaluation and draw conclusions and lessons learned that can be useful to other practitioners performing or studying confirmative evaluation.

Is Confirmative Evaluation Different from Formative or Summative Evaluation?

A significant underlying question in the literature is if confirmative evaluation is a distinct form of evaluation or if it is simply a different version of formative or summative evaluation. When Misanchuk (1978a) first described confirmative evaluation, they identified that confirmative evaluation shares similarities with both formative and summative evaluation. Because of these similarities, some researchers have assumed summative and confirmative evaluation are essentially the same thing, basing their determination off the belief that the only determining difference between the two is the evaluation's timing (Kang, 2012). This case illustrates the unique characteristics of confirmative evaluation that suggest it is a unique form of evaluation separate from formative and summative evaluation, but not for immediately obvious reasons.

Purpose

The key to identifying the differences among formative, summative, and confirmative evaluation is to focus on the different purposes of evaluation. For Van Tiem et al., (2000) confirmative evaluation focuses on value. This is inherently distinct in purpose from formative evaluation that focuses on creating improvement in a product and summative evaluation that focuses on proving improvement on a subject exposed to an intervention. As I argued in the

literature review, evaluation has four primary purposes: to improve a thing (formative), to determine if a thing fulfills the measure of its creation (summative), to determine the value or impact of a thing on its environment (confirmative), and to determine the degree to which effective evaluation has occurred (meta). This case helps clarify this concept. The evaluation undertaken in this case was intended to determine the degree to which the change to a new organizational model has been enacted. If the evaluation had been formative, its purpose would have been to improve the model or its implementation while it was being implemented. Had the evaluation been summative, it would have been to determine the degree to which the model works. In both cases, the questions and research methods would have changed based on the evaluation's purposes. Formative evaluation would have focused on questions like *What would make the new model or its implementation better?* and summative evaluation would have asked questions like *Does the new organizational model work?* These questions would have led to very different measures than were used in this case study. Instead, this case was guided by questions about to what degree the model had been implemented and what benefits stakeholders saw from it. From this purpose, it became apparent that both a time study and a review of stakeholder satisfaction were both necessary. The purpose of the evaluation shaped the questions asked and the approaches taken to answer those questions.

Timing

A commonly cited distinction of confirmative evaluation is its timing in an evaluative cycle. While I have argued that this is often misconceived as the primary determining factor of confirmative evaluation, its importance cannot be minimized. Misanchuk (1978c) noted that purpose and timing are interrelated. The purpose of an evaluation can establish when it should be conducted, and timing can influence the purpose of evaluation by influencing the questions being asked in the evaluation. In this case, confirmative evaluation was conducted over a span of time

four to eight months after the organizational change had been implemented. Had the confirmative evaluation been undertaken sooner, it is likely that there would have not been sufficient time to let the implemented changes take root and another round of confirmative evaluation would be needed to determine that the new organizational structure continues to operate as designed. Had the confirmative evaluation been undertaken a year after the organizational change, questions used may have been different based on the organization's impression and intention around its organizational structure and new vision for the organization. For example, during the Fall 2021 semester, the online learning organization in this case began considering logistics for using contracted international online instructors to teach some classes. This change to contracts and instructional staff would subsequently change the department's organizational model as affordances given in the current model would be untenable and potentially legally dangerous when operating with contracted international online instructors. While the results of this confirmative evaluation can be useful to the organization as it explores this new option, if the confirmative evaluation had been conducted a semester or two later the primary questions asked could have been very different and focused more on new relationships among remote leaders and contracted international instructors. In this way, timing informs the questions being asked in confirmative evaluation, but it does not overshadow the evaluation's purpose and its subsequent nature to identify the value or lasting impact of the organizational change.

Evaluator

Misanchuk (1978a) also argued that a difference among the types of evaluation is the person conducting the evaluation. They hypothesized that formative evaluation is conducted by someone with intimate knowledge of a product they are in the process of creating so they have

sufficient control to adjust the product while in development. They also said that summative evaluation should be conducted by someone with sufficient distance from the product to ensure sufficient objectivity. Misanchuk suggested that confirmative evaluation, because of its liminal nature between summative and formative evaluation, should be conducted by someone with intimate knowledge of the product or intervention but removed enough to be adequately objective. Because of this need for a balance between objectivity and control, Hellebrandt and Russell (1993) argued that it was likely best to have a team conduct confirmative evaluation.

In this case, the confirmative evaluation was conducted by a single instructor manager within the organization being studied. As an instructor manager, I had worked with the rest of the leadership team to design the new organization model as our managing director had envisioned. This involvement provided me with sufficient intimacy with the project to focus the confirmative evaluation on the aspects of the project that I knew would be potentially weak points in its implementation, namely how remote leaders implement their newly defined roles, while also providing me sufficient distance from the project because it was not *my* project in that I did not have a reputational stake in its success. If remote leaders failed to change how they functioned after receiving new job descriptions, the organizational change would have not occurred beyond name only and the organization would not realize the benefits of the change, but my job would continue to be secure. I had a small stake in the project, and without intimate knowledge of the prior model and the process taken to develop the new model, I likely would not have focused the confirmative evaluation on the time study of remote leaders.

While it may have made sense to have the entire instructor management team involved in the study so the confirmative evaluation was conducted by a team (Hellebrandt and Russell, 1993), doing so would also increase the necessary coordination and complexity of the study and

likely would not have improved the results of the time study or the review of extant data. Further, those conducting confirmative evaluation are faced with the challenge of securing sufficient buy in (Dessinger & Moseley, 2015) and overcoming the general tendency of organizations to expend more energy on adopting a change than implementing and following up on those changes (Fullan, 1996). Simply put, involving other instructor managers would have required convincing them to dedicate their already limited time to return to a project they had already completed and moved on from or had required members of other teams in the organization to be involved and likewise commit to the process. Had that occurred, the case could have potentially extended beyond the boundaries set upon it to examine just remote leaders' time in their roles, but it would have also increased the risk that the study would not be completed because of the project's competition for time with other initiatives. There is a cost and risk equilibrium in collaboration versus independent work to balance when conducting confirmative evaluation that impacts the quality and thoroughness of evaluation, its likelihood of being completed, and the how results are used that each instance of confirmative evaluation must weigh.

Tools

One area where confirmative evaluation is similar to both formative and summative evaluation are the tools all forms of evaluation use. Moseley and Solomon (1997) identified that confirmative evaluation uses multiple data-gathering instruments, including self-reporting, work-sample analysis, performance analysis, context studies, and cost-benefit comparisons to ask if customers' expectations are being met by the products they are provided. Likewise, formative and summative evaluation use similar tools. Atkin and Friemuth (2013) argued that formative evaluation use focus groups, secondary analysis, surveys, informal and formal feedback, extant

data, and interviews, all of which can be used in both summative and confirmative evaluation. Elwy et al. (2020) also argued that formative and summative evaluation used the same methods of research but differ in timing and purpose. Evaluation methods should produce information that is valid and useful within a specific decision-making context (Flagg, 2013), so the tools and methods used for evaluation are relatively agnostic and universally applicable if they are relevant to the questions driving the evaluation.

In this case, confirmative evaluation was conducted using time study methodology and surveys. While surveys are relatively ubiquitously utilized in different evaluation methods and models, time study methodology is an industrial or manufacturing measurement of worker productivity or used for worker movement improvement that would fall typically within a formative evaluation framework, it held relevance in this case as its benefits related to knowledge workers include improved job design, identification of redundancies, strategic planning, gap analysis, and establishing performance benchmarks (Ramírez & Nembhard, 2004). Other instances of confirmative evaluation will focus on different questions, and they will require equally contextually unique evaluation methods and tools. This case demonstrated that tools used in confirmative evaluation are not necessarily unique to confirmative evaluation. Like the evaluator, tools used in confirmative evaluation can add to the confusion around what is and is not confirmative evaluation because they are also used in other forms of evaluation. Tools used in confirmative evaluation are not necessarily unique; the thread that detangles confirmative evaluation is again the purpose for which those evaluative tools are used, which again, is closely intertwined but not solely reliant on the timing of the evaluation.

Is Confirmative Evaluation Real?

From this case and its subsequent analysis, it becomes clear that the defining factor of what makes confirmative evaluation different from formative, summative, and meta evaluation is not the evaluator, the tools, or timing of the evaluation. Rather, the purpose behind the evaluation defines the type of evaluation undertaken, which then informs the timing and other characteristics of the evaluation. By focusing on purpose first, we see that confirmative evaluation plays a distinct role in an evaluation system, even while it can be conducted by the same people who conduct formative or summative evaluation with the same tools they would use for formative or summative evaluation. An evaluation's timing is informed by its purpose, but it does not solely define the type of evaluation undertaken and should be noted after the purpose of an evaluation, not as its defining feature.

What Role Does Confirmative Evaluation Play in Performance Improvement and Continuous Improvement Models?

Having established that confirmative evaluation is a distinct form of evaluation with a distinct purpose but without necessarily distinct methods or practitioners, the next question about confirmative evaluation addresses its role in performance improvement and continuous improvement processes. Where does it fit in our different performance improvement models?

While Misanchuk (1978) argued that confirmative evaluation added to the formative/summative dichotomy to make an evaluation trichotomy, Hellebrandt and Russell (1993) extended their argument and stated that confirmative evaluation makes the trichotomy a cycle of evaluation. They showed through their hypothetical examples that confirmative evaluation, because of its purpose and timing, allows organizations that use it to make decisions about interventions after they have been implemented that can help them intelligently continue,

revise, or end the implementation. If the implementation is ended, no further evaluation is needed. If the implementation is continued without modification, additional confirmative evaluation is warranted later. If the implementation is revised, the implementation moves into redevelopment and subsequent formative evaluation, to kick off the cycle again.

In this case, confirmative evaluation was used to identify the degree to which remote leaders' work aligns with their job descriptions and to what degree stakeholders are satisfied with that work since the new organizational model and remote leader job descriptions had been implemented. Had the results from the time study shown vast discrepancy in the time spent versus the expected time represented by the job descriptions of remote leaders, or had stakeholders been grossly dissatisfied with remote leaders since adopting their new job descriptions, the organization could have discussed if the new organizational model should be abandoned. Because the results of this example of confirmative evaluation showed overall satisfaction along with specific continued gaps in performance, the organization could move into another round of refinement in remote leaders' work, if the coming cause analysis revealed the cost of the gaps warranted further action (Kaufman, 1996). This case demonstrated the potential for confirmative evaluation to restart the continuous improvement cycle where it might have otherwise ended.

Moseley & Hasting's (2005) Intervention Implementation Process Model

One model this can be applied is Moseley and Hasting's (2005) four-stage Intervention Implementation Process Model. Moseley and Hastings (2005) argued that the Plan, Do, Stabilize, Institutionalize change management process is a cyclical one that acts as a roadmap for performance improvement practitioners to work through the intricacies of managing interventions. They argued that each stage is marked by four key actions: communication, action,

auditing, feedback. In their explanation of the model, formative evaluation occurs in the Plan, Do, and Stabilize stages while summative evaluation occurs in the Institutionalize stage.

According to their definition, the Institutionalize stage is unique from the other stages because it determines the success of the intervention, how closely it aligns with the organization's mission, values, and beliefs, and what barriers to success remain. They also recognize that the feedback from this stage serves as input for Stage 1 for future continuous improvement efforts while simultaneously laying the foundation for confirmative evaluation.

There are two issues with this explanation as it relates to confirmative evaluation. First, this model intentionally removes confirmative evaluation from the continuous improvement cycle, which perhaps unintentionally suggests that it lacks value in that continuous improvement cycle. If confirmative evaluation focuses on questions related to an intervention's merit, worth, value, or impact, it is then immensely relevant within the cycle because when an intervention is found to not be worthwhile to an organization its continued existence should be questioned and potentially ended rather than continuously improved. Also problematic to removing confirmative evaluation from a continuous improvement cycle is that confirmative evaluation naturally aligns with the purpose of the institutionalization stage in this model. Institutionalization is meant to resolve lingering issues, which can be a direct effect of confirmative evaluation in continuous improvement. Additionally, Institutionalization is an opportunity to reaffirm the organization's long-term commitment to the change and requires upper management's clear support (Moseley & Hastings, 2005), which are also hallmarks of confirmative evaluation because it requires allocating resources after an intervention has been adopted (Dessinger & Moseley, 2015). Confirmative evaluation naturally aligns with the institutionalization stage of this model and

belongs in the continuous improvement cycle as a vital element of it, not outside as a potential add on.

Second, this model misnames the evaluation that occurs in the Institutionalization stage. Moseley and Hasting (2005) argued that summative evaluation in the Institutionalization stage focuses on analyzing data to determine how successful the intervention is at closing the gap and how closely aligned the intervention is with organizational mission, beliefs, and values. The problem with the stance that summative evaluation does this in the institutionalization stage is that they also claim it does the same thing during the Stabilization stage, but the purposes of each stage are inherently different and thus require a different focus of evaluation. Further, the Do stage focuses auditing on assessing the impact of an intervention on individuals and organizations, which is what summative evaluation is about. Having already established how well an intervention closes the gap it was intended to close in the Stabilize stage, a better use of the Institutionalize stage is to ensure the intervention has stuck over time and is producing the desired effects. This is precisely what confirmative evaluation does.

By focusing on Moseley and Hasting's (2005) Intervention Implementation Process Model through the evaluative purposes and trichotomy, we see that Formative, Summative, and Confirmative evaluation are all actually present in this model (see Figure 8). By using the Evaluation Conceptualization Framework presented in my literature review to analyze Moseley and Hasting's (2005) Intervention Implementation Process Model (see Table 21), we see that the types of evaluations change throughout the process depending on the purposes of evaluation present in each of the four stages. At some points, formative and summative evaluation occur within the same stage, highlighting the differences in evaluation are more a product of their purposes than their timing. Interestingly, but not central to this study, we also see that meta

evaluation occurs in this model throughout the cycle as a means of moving lessons learned throughout an implementation to useful knowledge in subsequent implementations.

Figure 8

Moseley & Hastings Intervention Implementation Process Model, Revised

Moseley & Hasting’s (2005) Intervention Implementation Process Model, Revised

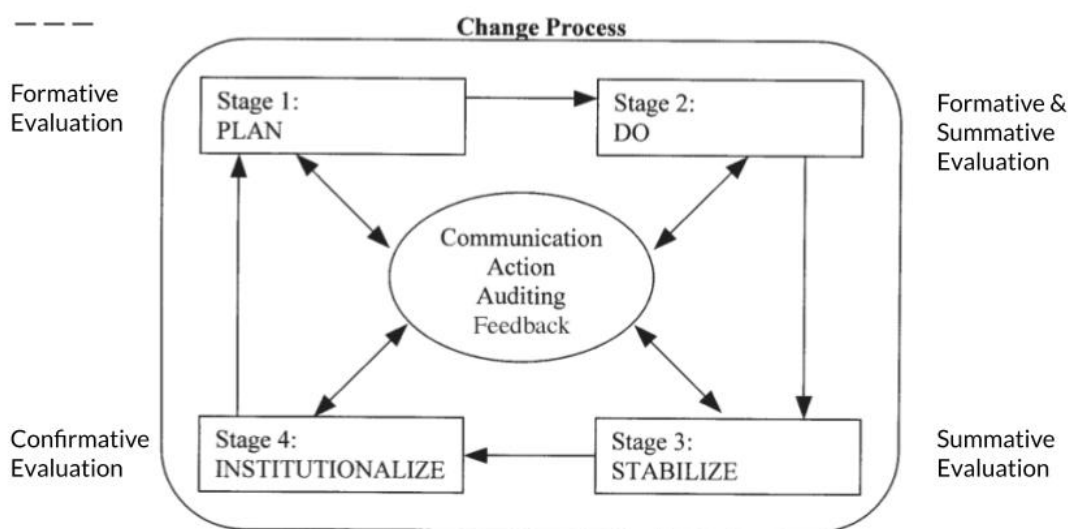


Table 22

Evaluation Conceptualization Framework Applied

Evaluation Type <i>Purpose</i>	Descriptors	Model
		Moseley & Hasting’s (2005) Intervention Implementation Process Model
Formative <i>To improve a thing</i>	Timing	In implementation processes Stage 1: Plan Stage 2: Do
	Sources	Surveys, interviews, focus groups, observations, extant data reviews

Table 21 continued

Evaluation Type <i>Purpose</i>	Descriptors	Model
		Moseley & Hasting's (2005) Intervention Implementation Process Model
Formative <i>To improve a thing</i>	Outputs	Identify and remove barriers to success, inputs for subsequent stages, input for future implementations
	Etc.	
Summative <i>To determine if a thing fulfills the measure of its creation</i>	Timing	In implementation processes Stage 2: Do Stage 3: Stabilize
	Sources	Unspecified
	Outputs	Assessment of impact of intervention on individuals and the organization, review of success in completing launch tasks, measurement of comfort of individuals with the change, identification of barriers to success, identification and validation of success of the process and its champions, evaluation of the intervention's success in closing the intended gap, preparation to move to Stage 4
	Etc.	
Confirmative <i>To determine the value or impact of a thing</i>	Timing	In implementation processes Stage 4: Institutionalize
	Sources	Unspecified
	Outputs	Data that summarizes the degree to which the intervention closes the performance gap and to what degree the intervention aligns with organizational mission values, and beliefs; identification of additional performance problems or opportunities; input for Stage 1 of future implementations; long term stabilization
	Etc.	
Meta <i>To determine the degree to which effective evaluation has occurred</i>	Timing	In implementation processes Stage 1: Plan Stage 2: Do Stage 3: Stabilize Stage 4: Institutionalize
	Sources	Feedback from each stage, methods unspecified
	Outputs	Identification and documentation of organizational barriers to successful implementation, modified processes for current and future implementations
	Etc.	

From this example of Moseley and Hasting's (2005) Intervention Implementation Process Model, we see confirmative evaluation plays an integral role in the implementation of interventions with people and organizations. Rather than being an evaluation completely distinct from the rest of the process, confirmative evaluation completes the evaluation cycle and helps solidify interventions by promoting upper management buy-in, identifying the degree to which an intervention really closes a gap, and sets the stage for future implementations. Confirmative evaluation also ensures the intervention is aligned with organizational mission, values, and beliefs. When the case in this study is considered through the lens of confirmative evaluation in Moseley and Hasting's (2005) Intervention Implementation Process Model, it has served these purposes by showing the organization the degree to which remote leaders are bought into their roles and performing them as designed and the effects of their work on their stakeholders. Its list of identified continuing gaps in performance also lays the foundation for subsequent intervention implementations.

Marker et al.'s (2014) Spiral HPI Framework

Another model in which this case helps refine our use of understanding of confirmative evaluation in the continuous improvement cycle is Marker et al.'s (2014) Spiral HPI Framework. Hastings (2009) argued that the problem with the formative, summative, and confirmative terminology is that it assumes each must occur within a specific timeframe, and that each type of evaluation is relevant throughout the performance improvement cycle. They argued that all four forms of evaluation should be addressed in all phases of the HPT Model (Van Tiem et al., 2012), which is exactly what the ISPI HPT model demonstrates by tying each stage to evaluation generally rather than a single type of evaluation. This design emphasizes that performance improvement practitioners must consider and plan in all four types of evaluation where it is

relevant to do so, giving them flexibility to apply the model as context specifies. But a problem with this model is that continues to suggest interventions happen linearly and is not reflective of how performance improvement works in the real world (Marker et al., 2014; Czeropski & Pembroke, 2017).

Marker et al.'s (2014) Spiral HPI Framework provides an alternative to Van Tiem et al.'s (2012) HPT model to better illustrate the fluid, iterative nature of HPI work and better demonstrates a continuous performance improvement model than the ISPI HPT model. It also acknowledges that continuous improvement is often undertaken in a rapid prototyping style rather than a step-by-step process suggested by the ISPI HPT model. A challenge with the Spiral HPI Framework is that it ends evaluation at the same time as the implementation and change facilitation. In practice, doing so limits practitioners' ability to determine the real value of a change to an organization's valued outcomes. If we are focused only on what will improve an intervention and ensuring that it functions as it ought, this is sufficient. If we want to determine the impact of an intervention, evaluation needs to extend into the maintenance stage of an intervention to ensure it continues to do what it was intended to and that it aligns with the organization's vision, mission, and goals (see Figure 2). Confirmative evaluation is what can do this.

This case demonstrated this value of confirmative evaluation in a revised Spiral HPI Framework. Chronologically, the case occurred after the intervention of changing the organizational structure had been implemented and the new organizational structure was in the maintenance phase. It identified areas where the implementation had been successful and where it needed continued effort. Additionally, this case addressed valued outcomes for the organization: course lead satisfaction with ACL performance and instructor satisfaction in the

new organization. Interestingly, the results of this instance of confirmative evaluation also identified where continued gaps exist with the new organizational structure, causing the process to wrap back around to analysis and restart the process again. Marker's model is not meant to show a continuous improvement process and rather focuses on the idea that practitioners work on many components or process simultaneously and move from one to another as dictated by need and context, so it is difficult to visualize in this model how introducing confirmative evaluation into it works the process back to analysis and restarts the spiral iterations through the intervention selection, design, development, implantation, and change processes.

The spiral iterations in this model also present a problem for confirmative evaluation in continuous improvement because the spirals emphasize the value of rapid prototyping. If the value of rapid prototyping is that it is *rapid* or that iterations are applied quickly and frequently (Marker et al., 2014), inserting confirmative evaluation into every cycle or iteration as might be assumed by my earlier argument that confirmative evaluation completes the evaluation cycle would inevitably slow the prototyping and be counterproductive. This is what Finney (2020) argued for in their inclusion of confirmative evaluation immediately after each step of the CIPP model as a tie to the next step. The problem with this is that because confirmative evaluation's purpose and its timing are perhaps inextricably interconnected, inserting confirmative evaluation into each iteration drastically slows the iterative process. Because the purpose of confirmative evaluation is to determine if an intervention continues to serve its purpose or if it presents value to the customer, inserting confirmative evaluation at some phase in the model is useful, but not at every iteration. Marker et al. (2014) argued that formative, summative, confirmative, and meta evaluation all can occur throughout the process, but in this model it appears that confirmative evaluation would be counterproductive if applied *throughout* the model.

How confirmative evaluation is applied in this model depends largely on what is being iterated and the lifespan of the process. Relatively small interventions may be completed in a shorter timeframe and might not have confirmative evaluation applied to them until after the intervention has been in place for some time, as is typically addressed in the literature (Dessinger & Moseley, 2015). In this way, the revised Spiral HPI Framework (see Figure 2) I presented in Chapter 2 is how we would see confirmative evaluation add value to the iterative process. It can be used after rapid prototyping to examine the results of the iterations and provide feedback to inform future iterations or interventions.

But what about projects with longer lifespans and longer gaps of time between iterations? Not all projects' iterations are "rapid," while still following the spiral iterations presented in Marker's model. These projects develop more slowly and over greater lengths of time but are still iterative in nature. In these cases, we may see confirmative evaluation interspersed throughout the iterations at strategic moments where the output of the confirmative evaluation can add another spiral of iteration that would otherwise not be developed. The case in the current study demonstrates this. The time study of remote leaders' work and the review of stakeholder satisfaction produced data that could guide the next iteration of changes to the organizational model or training of remote leaders if the organization determined that the gaps presented in this case warranted action. In this way, confirmative evaluation helps determine if the iterations continue or end. If iterations end, confirmative evaluation appears as presented in Figure 2, but if iterations continue, it will appear somewhere or possibly in many places along the evaluation component of this model. The formative, summative, confirmative cycle will happen until an organization determines that the intervention no longer fits its valued outcomes. Confirmative evaluation is what can inform that decision.

What Challenges Are Faced when Conducting Confirmative Evaluation?

Identifying where confirmative evaluation fits in different continuous improvement models raises the next question of how this case demonstrated the challenges that are faced while conducting confirmative evaluation. These challenges were limitations in this study that can be addressed in future cases of confirmative evaluation.

Including confirmative evaluation into the continuous improvement process can be difficult because continuous improvement implies that there is no end to the cycle. Confirmative evaluation forces organizations to go back to previous interventions and see if they are working when they would rather keep moving to the next intervention. The value in this is that it can strengthen future interventions, but it is a hard sell to get organizations to stop and reflect on what they have done, especially when doing so appears to be costly in time and resources. The literature about confirmative evaluation suggested that timing and conceptual obstacles (Dessinger & Moseley, 2015), organizational barriers (Dessinger & Moseley, 2015; Marshall & Rossett, 2014), tendencies for organizations to expend more energy in adopting change than in implementing it (Fullan, 1996), the relative cost of confirmative evaluation (Jackson, 1989; Guerra-Lopez, 2008), the lack of incentives to conduct confirmative evaluation (Marshall & Rossett, 2014), and a lack of understanding about confirmative evaluation (DeVaughn & Stefaniak, 2020b) all hinder the use of confirmative evaluation. This case provided clearer understanding of these issues.

Timing

Dessinger and Moseley (2015) argued that a challenge to confirmative evaluation is that because it is conducted over relatively long periods of time, original measures produced at the beginning of an intervention lack flexibility and monitoring and may not be applicable to the

confirmative evaluation needed after an intervention has been implemented. In this case, confirmative evaluation was not planned into the change process, so measures that led to adopting the new organizational model were unavailable or unapplicable to the new questions being asked. Had confirmative evaluation been considered during the development of the implementation plan for the new organization model and subsequently revised job descriptions for remote leaders, measures could have been developed that would have had more applicability after the intervention and would have provided a stronger comparison between stakeholders satisfaction with the new remote leaders before and after the new model was instituted, This case highlights the need to consider from the beginning how interventions will be confirmed so stronger measures can be implemented.

Conceptual Barriers

Dessinger and Moseley (2015) also argued that because confirmative evaluation seeks to determine the value or impact of an intervention, it often relies on complex formulas and data. This case eschewed this issue by narrowing the scope of the evaluation, which simplified the data collection. A limitation in this study was that it did not go so far as to place a dollar value on the organizational changes and subsequent stakeholder satisfaction as Phillips et al.'s (2006) ROI model or Kirkpatrick's Business Partnership Model (2009). But in this case, neither of these two models were necessary because the organization was satisfied knowing that the organizational changes decreased operating expense and was not a barrier or driving factor in adopting these changes. Because of this I was able to limit the scope of this case to job performance and stakeholder satisfaction, which made the actual research in the case more practical and simpler. This demonstrates that under the right circumstances, confirmative evaluation does not necessarily require complex formulas and data, and that the questions asked in confirmative

evaluation are contextually relevant. In future research of confirmative evaluation, it will be useful to identify organizational values to select the appropriate confirmative evaluation measures.

Organizational Barriers

Marshall and Rossett (2014) identified that organizational barriers to evaluation generally inhibit designers' ability to evaluate the impact of their interventions. Organizational barriers include environmental factors, lack of incentives, lack of tools and systems for evaluation and reporting, and lack of support from more skilled data analysts. In this case, the organization and the research did not lack tools, systems, or the support of skilled data analysts since the organization has a robust institutional research team that was available for support while developing and implementing the time study and using the previously developed Course Group and Course Council surveys. Survey systems were already in place, which facilitated this case of confirmative evaluation. The challenges that *were* faced in this case were environmental factors and lack of incentives. Because remote workers work across the United States and because they perform their work at times of the day and week that fit their personal schedules, a traditional time study methodology could not be used as they typically require physical proximity and direct observation of workers (Yazdi et al., 2019). This challenge was addressed through self-reporting of time performed in remote leaders' duties, which was then triangulated against observational data from the Course Council and Course Group surveys and found to be consistent across these measures. Had these two surveys not been available to use or shown significant difference in what stakeholders observed of remote leaders compared to what they reported of themselves, the trustworthiness of the time study data would have been compromised. Future researchers of confirmative evaluation would do well to identify how extant data can be used to help bolster the

legitimacy of the data collected during confirmative evaluation, especially when there are not direct measures available from before the intervention is implemented. Again, considering how an intervention might be evaluated over its lifecycle during the inception of its development will substantially aid in confirmative evaluation efforts later and provide decision makers with the data needed to continue, revise, or end the initiative.

Adoption Versus Implementation Costs

Part of the challenge performance improvement practitioners face when conducting confirmative evaluation stems from the tendency of organizations to expend more energy adopting a change than implementing it (Fullan, 1996), resulting in decisions being made without sufficient follow up. Because of this and not designing it into the intervention's implementation, confirmative evaluation becomes a costly endeavor (Jackson, 1989; Guerra-Lopez, 2008). In this case, this issue was addressed by relying on existing remote leaders and their managers to conduct the time study. No additional staff were needed to conduct the research necessary, and the time required to manage the deployment of surveys fit within the researcher's time without interfering with their regular duties after the researcher developed the process and measures for this research. Further, having an instructor manager in charge of this confirmative evaluation limited the time required by other instructor managers and allowed them to focus their energies on other projects. The biggest potential cost for the time study was the time it would take remote leaders to track their time over the semester each week, which is why the matrix sampling methodology (Thorndike & Thorndike-Christ, 2010) was selected for this case. This study demonstrated that the cost of confirmative evaluation can be mitigated through combining new research with extant data, can provide a deeper analysis of a training program or intervention's total impact, and can lead to further innovation and performance improvement.

Incentives

Matrix sampling reduced the number of required participants each week of the time study and the subsequent draw on remote leaders' time, but it did not address the lack of incentives for remote leaders to participate in the data collection for this case (Marshall & Rossett, 2014). In this case, remote leaders were informed of the purpose for conducting the time study and were invited but not required to participate in the time study. There was no direct benefit to participants in the study, and there were perceived risks in participating. Some participants emailed the researcher during the time study to let them know that the week they were asked to report their time was not indicative of the weeks in which they were most active, suggesting concern that the time study would reflect poorly on themselves in their remote leader functions. This was addressed by repeatedly explaining that their responses were anonymous and could not be tied to any single respondent, but the concern was still raised occasionally throughout the study. In addition, remote leaders could officially opt out of the study or simply choose to not complete the surveys used to collect data, which reduced the participation rates to an average of 37% each week. Higher participation rates would have provided more trustworthy results. Future cases of confirmative evaluation that rely on survey responses would do well to consider incentives to help improve response rates.

Conclusion & implications for future study

Even with the challenges and limitations in this case, this instrumental case study illustrated several useful insights about confirmative evaluation. First, it demonstrated that confirmative evaluation is a distinct form of evaluation because its purpose is distinct from formative and summative evaluation. The most common approach to distinguishing these three types of evaluation focuses on their timing. While this does draw boundaries between these three

evaluations, it is less distinctive than their purposes. The purpose of evaluation dictates the timing of the evaluation, and therefore is a stronger definitive descriptor of each type of evaluation than timing alone.

This study also demonstrated that confirmative evaluations have a distinct role in continuous improvement. As part of the continuous improvement process rather than an afterthought to it, confirmative evaluation can help institutionalize interventions and extend the continuous improvement cycle beyond its initial implementation. These effects can help the intervention address the organization's valued outcomes. In this process, confirmative evaluations helped highlight the value of the intervention to the organization, which is something other forms of evaluation cannot do.

Finally, this case helps address a gap in our literature by highlighting one real life example of confirmative evaluation in action. Our field's ability to use confirmative evaluation is hampered by a dearth of examples of confirmative evaluation in the real world. As we see confirmative evaluation in action in unique cases, we can come to understand its real value in continuous improvement processes and to those who design and implement instructional or non-instructional interventions. Currently, we lack sufficient knowledge about confirmative evaluation's impact and how to do it. Those who train future instructional designers lack examples and personal experience in confirmative evaluation (DeVaughn & Stefaniak, 2020b), so future designers are also uninformed of this tool that can prove to be a valuable asset to designers, especially when they need to demonstrate the value of their work. By examining real instances of confirmative evaluation, we can come to understand this unique form of evaluation and develop a stronger ability to conduct it. This case showed confirmative evaluation in an online higher education context; cases that examine confirmative evaluation in other professional

or business contexts will help show confirmative evaluation's utility and adaptability. This case showed confirmative evaluation within a narrow scope and ended before its total effects could be seen; cases that are not bounded like this case and that can examine confirmative evaluation in the context of the entire continuous improvement cycle will help demonstrate confirmative evaluation's role the continuous improvement process. This case showed confirmative evaluation implemented without initially being planned into the intervention; cases that show its utility when planned in from the beginning would show its value to an organization throughout the entire process and would likely offer low-cost measures to support Confirmative Evaluation.

The assumption that confirmative evaluation is costly remains a challenge to its adoption. This case showed confirmative evaluation conducted with little additional cost to the organization but with limited scope on one aspect of the organizational change; other cases that show confirmative evaluation conducted in low cost-high value ways would be incredibly helpful for practitioners and those that train future IDs and performance improvement professionals. Equally useful would be more cases that demonstrate how confirmative evaluation can be undertaken to provide other unique methods that can help practitioners consider contextually useful approaches to confirmative evaluation. Doing so would provide designers with a more robust toolbox for future endeavors.

Confirmative evaluation is a useful but underutilized tool for instructional designers and performance improvement professionals. This case has been instrumental in better understanding confirmative evaluation, but additional cases will help our field to develop a more mature understanding of confirmative evaluation and stronger strategies for how to effectively use it.

REFERENCES

- Adams, C. D., Baaki, J., & Stefaniak, J. E. (2020). Challenges faced by certified performance technologists in conducting needs assessment. *Performance Improvement Quarterly*, 33(4), 419-442. DOI: 10.1002/piq.21329
- Alarifi, S. A., & Alamri, A. (2014). HPT and six sigma: Is there a difference that matters? *Performance Improvement*, 53(7), 14-22. DOI: 10.1002/pfi.21412
- Antony, J., Snee, R., & Hoerl, R. (2017). Lean six sigma: Yesterday, today and tomorrow. *International Journal of Quality & Reliability Management* 34(7), 1073-1093. DOI: 10.1108/IJQRM-03-2016-0035
- Ashraf, J. (2017). Examining the public sector recruitment and selection, in relation to job analysis in Pakistan. *Cogent Social Sciences*, 3(1), 1309134. DOI: doi.org/10.1080/23311886.2017.1309134
- Atkin, C. K., & Freimuth, V. (2013). Guidelines for formative evaluation research in campaign design. In R. E. Rice & C. K. Atkin (Eds.), *Public communication campaigns* (4th ed., pp. 53-68).
- Barbouletos, S. (2011). Discrepancy between role expectations and job descriptions: the impact on stress and job satisfaction. [Master's thesis, University of Washington].
[https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/17087/Barbouletos-Capstone%20\(2\).pdf;sequence=1](https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/17087/Barbouletos-Capstone%20(2).pdf;sequence=1)
- Baskarada, S. (2014). Qualitative case study guidelines. *The Qualitative Report*, 19(40), 1-25.
- Beaudoin, M. (2016). Issues in distance education: A primer for higher education decision makers. *New Directions for Higher Education*, 2016 (173), 9-19. <https://doi.org/10.1002/he.20175>

- Binder, C. (2001). Measurement: A few important ideas. *Performance Improvement*, 40(3), 20-28.
- Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations*. Jossey-Bass.
- Boulding, K. E. (1956). General systems theory—the skeleton of science. *Management Science*, 2(3), 197-208.
- Brigance, S. K. (2011). Leadership in online learning in higher education: Why instructional designers for online learning should lead the way. *Performance Improvement*, 50(10), 43-48.
- Carliner, S., Castonguay, C., Sheepy, E., Ribeiro, O., Sabri, H., Saylor, C., & Valle, A. (2015). The job of a performance consultant: a qualitative content analysis of job descriptions. *European Journal of Training and Development*, 39(6), 458-483. DOI 10.1108/EJTD-01-2015-0006
- Chen, H. T. (1996). A comprehensive typology for program evaluation. *Evaluation Practice*, 17(2), 121-130. DOI: 10.1177/109821409601700204
- Childs, R. A., & Jaciw, A. P. (2002). Matrix sampling of items in large-scale assessments. *Practical Assessment, Research, and Evaluation*, 8(1), DOI: <https://doi.org/10.7275/gwvh-4z51>
- Christensen, C. M., & Eyring, H. J. (2011a). *The innovative university: Changing the DNA of higher education from the inside out*. John Wiley & Sons.
- Christensen, C. M., & Eyring, H. J. (2011b). How disruptive innovation is remaking the university. *Harvard Business School Newsletter*, 25.
- Czeropski, S., & Pembroke, C. (2017). E□ Learning ain't performance: Reviving HPT in an era of agile and lean. *Performance Improvement*, 56(8), 37-45. DOI: 10.1002/pfi.21728

- Dixon, N. M. (1990). *Evaluation: A tool for improving HRD quality*. University Associates. Inc.
- Dessinger, J. C., & Moseley, J. L. (2015). *Confirmative evaluation: Practical strategies for valuing continuous improvement*. John Wiley & Sons.
- Dessinger, J. C., Moseley, J. L., & Van Tiem, D. M. (2012). Performance improvement/HPT model: Guiding the process. *Performance Improvement*, 51(3), 10-17. DOI: 10.1002/pfi.20251
- DeVaughn, P., & Stefaniak, J. (2020a). An exploration of the challenges instructional designers encounter while conducting evaluations. *Performance Improvement Quarterly*, 33(4), 443-470. DOI: 10.1002/piq.21332
- DeVaughn, P., & Stefaniak, J. (2020b). An exploration of how learning design and educational technology programs prepare instructional designers to evaluate in practice. *Educational Technology Research and Development*, 68(6), 3299-3326. DOI: 10.1007/s11423-020-09823-z
- Draper, F. J. (2012). Development of a student-centered evaluation framework for environmental vocational education and training courses (Doctoral dissertation). Retrieved from <https://bradscholars.brad.ac.uk/handle/10454/5496>
- Ellsworth, J. B. (1997). Technology and change for the information age (ED439702). ERIC. <https://files.eric.ed.gov/fulltext/ED439702.pdf>
- Elwy, A. R., Wasan, A. D., Gillman, A. G., Johnston, K. L., Dodds, N., McFarland, C., & Greco, C. M. (2020). Using formative evaluation methods to improve clinical implementation efforts: description and an example. *Psychiatry Research*, 283, 1-6. doi.org/10.1016/j.psychres.2019.112532

- Finney, T. L. (2020). Confirmative Evaluation: New CIPP evaluation model. *Journal of Modern Applied Statistical Methods*, 18(2), 30-54. DOI:, eP3568.
- Flagg, B. N. (2013). *Formative evaluation for educational technologies*. Routledge.
- Fullan, M. (1996). Implementation of innovations. In M. Eraut (ed.), *The International encyclopedia of educational technology* (pp. 273-281). Elsevier.
- Giberson, T. R. (2015). Industrial–organizational psychology and the practice of performance improvement. *Performance Improvement Quarterly*, 28(2), 7-26. DOI: 10.1002/piq.21191
- Giberson, T. R., Tracey, M. W., & Harris, M. T. (2006). Confirmative evaluation of training outcomes. *Performance Improvement Quarterly*, 19(4), 43-61. doi: 10.1111/j.1937-8327.2006.tb00384.x
- Gonzalez, J. M., & Eltinge, J. L. (2007, July). Multiple matrix sampling: A review. In *JSM Proceedings, Survey Research Methods Section* (pp. 3069-3075).
- Griffin, R. (2012). A practitioner friendly and scientifically robust training evaluation approach. *Journal of Workplace Learning* 24(6), 393-402. DOI 10.1108/13665621211250298
- Guerra-Lopez, I. J. (2008). *Performance evaluation: Proven approaches for improving program and organizational performance* (vol. 21). John Wiley & Sons.
- Harling, K. (2002, July 27). Case studies: Their future role in agricultural and resource economics [Conference presentation abstract]. The annual meeting of the American Agricultural Economics Association, Long Beach, California, United States.

- Hastings, N. B. (2009). Integrated evaluation: Improving performance improvement. *Handbook of improving performance in the workplace: Volumes 1-3*, 240-255. DOI: 10.1002/9780470592663.ch67
- Heinich, R. (1985). Instructional technology and the structure of education. *Educational Communications and Technology Journal*, 33(1), 9-15. DOI: 10.1007/BF02769337
- Hellebrandt, J., & Russell, J. D. (1993). Confirmative Evaluation of Instructional Materials and Learners. *Performance and Instruction*, 32(6), 22-27. DOI: 10.1002/pfi.4170320608
- Imai, M. (1997). *Gemba kaizen: A commonsense, low-cost approach to management*. McGraw-Hill
- Jackson, T. (1989). *Evaluation: Relating training to business performance*. Pfeiffer & Company.
- Kalman, H. K. (2016). Integrating evaluation and needs assessment: A case study of an ergonomics program. *Performance Improvement Quarterly*, 29(1), 51-69. DOI: 10.1002/piq.21214
- Kang, S. P. (2012). *Validation of key stages of the international society for performance improvement human performance technology model* (Doctoral dissertation). Retrieved from Proquest. (3498934)
- Kaplan, D., & Su, D. (2016). On matrix sampling and imputation of context questionnaires with implications for the generation of plausible values in large-scale assessments. *Journal of Educational and Behavioral Statistics*, 41(1), 57-80. DOI: 10.3102/1076998615622221
- Karger, D. W., & Bayha, F. H. (1987). *Engineered work measurement: The principles, techniques, and data of methods-time measurement background and foundations of work measurement and methods-time measurement, plus other related material*. Industrial Press Inc.

- Kaufman, R. (1996). What works and what doesn't: Evaluation beyond Kirkpatrick. *Performance and Instruction*, 35(2), 8-12. DOI: 10.1002/pfi.4170350204
- Kennedy, P. E., Chyung, S. Y., Winiiecki, D. J., & Brinkerhoff, R. O. (2014). Training professionals' usage and understanding of Kirkpatrick's level 3 and level 4 evaluations. *International Journal of Training and Development*, 18(1), 1-21. DOI: 10.1111/ijtd.12023
- Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating training programs: The four levels*. Berrett-Koehler Publishers.
- Kirkpatrick, J. (2009). The Kirkpatrick model: Past, present and future. *Chief Learning Officer*, 8(11), 20-55.
- Kshatriya, S. (2016). Linkage of job analysis to performance management: a case study. *International Journal of Management, IT and Engineering*, 6(9), 202-222.
- Klassen, K. J., Russell, R. M., & Chrisman, J. J. (1998). Efficiency and productivity measures for high contact services. *Service Industries Journal*, 18(4), 1-18. DOI: 10.1080/02642069800000038
- Kotter, J. (1995). Leading change: why transformation efforts fail. *Harvard Business Review*, 73(2), 55-67.
- Kulkarni, P. P., Kshire, S. S., & Chandratre, K. V. (2014). Productivity improvement through lean deployment & work study methods. *International Journal of Research in Engineering and Technology*, 3(2), 429-434. DOI: IJRET_110302076
- Langdon, D. (2000). *Aligning performance: Improving people, systems, and organizations*. Jossey-Bass/Pfeiffer.
- Laszlo, E. (1996). *The systems view of the world: A holistic vision for our time*. Hampton Press.

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, 2(4), 34-46.

<https://doi.org/10.1111/j.1540-4560.1946.tb02295.x>

Mader-Clark, M. (2013). *The job description handbook*. Nolo.

Marker, A., Villachica, S.W., Stepich, D., Allen, D., & Stanton, L. (2014). An updated framework for human performance improvement in the workplace: The spiral HPI framework. *Performance Improvement*, 53(1), 10–23. DOI: 10.1002/pfi.21389

Marshall, J., & Rossett, A. (2014). Perceptions of barriers to the evaluation of workplace learning programs. *Performance Improvement Quarterly*, 27(3), 7-26. DOI: 10.1002/piq.21173

Menon, S., & Suresh, M. (2020). Factors influencing organizational agility in higher education. *Benchmarking: An International Journal*, 28(1), 307-332.

Mills, A. J., Durepos, G., & Wiebe, E. (2010). Instrumental case study. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.). *Encyclopedia of case study research* (pp. 474-475). Sage.

Misanchuk, E. R. (1978a). Uses and abuses of evaluation in continuing education programs: On the frequent futility of formative, summative, and justificative evaluation (ED160734). ERIC. <https://eric.ed.gov/?id=ED160734>

Misanchuk, E. R. (1978b). Evaluation types in instructional development: Beyond the formative-summative distinction (ED153646). ERIC. <https://eric.ed.gov/?id=ED153646>

Misanchuk, E. R. (1978c). Descriptors of evaluations in instructional development: beyond the formative-summative distinction. *Journal of Instructional Development*, 2(1), 15-19. DOI: 10.1007/BF02904322

- Mohamed, F. R., & Hossny, E. K. (2020). Role Clarity as a Predictor of Nurses' Job Satisfaction. *Assiut Scientific Nursing Journal*, 8(21), 65-73. DOI: 10.21608/asnj.2020.29533.1016
- Moldovan, L. (2016). Training outcome evaluation model. *Procedia Technology*, 22, 1184-1190. DOI: 10.1016/j.protcy.2016.01.166
- Moseley, J. L., & Hastings, N. B. (2005). Implementation: The forgotten link on the intervention chain. *Performance Improvement*, 44(4), 8-14. DOI: 10.1002/pfi.4140440405
- Moseley, J. L., & Solomon, D. L. (1997). Confirmative evaluation: A new paradigm for continuous improvement. *Performance Improvement*, 36(5), 12-16. DOI: 10.1002/pfi.4140360504
- Overby, W. M. (1983). Technique for group time measurement simplifies indirect labor observations. *Industrial Engineering*, 15(7), 34-40.
- Passmore, J., & Velez, M. (2012). SOAP□M: A training evaluation model for HR. *Industrial and Commercial Training*, 44(6), 315-325. DOI 10.1108/00197851211254743
- Pató, B. S. G. (2017). Formal options for job descriptions: Theory meets practice. *Journal of Management Development*, 36(8), 1008–1028. <https://doi.org/10.1108/JMD-01-2016-0019>
- Patton, M. Q. (1996). A world larger than formative and summative. *Evaluation Practice*, 17(2), 131-144. DOI: 10.1177/109821409601700205
- Perez-Soltero, A., Aguilar-Bernal, C., Barcelo-Valenzuela, M., Sanchez-Schmitz, G., Meroño-Cerdan, A. L., & Fornes-Rivera, R. D. (2019). Knowledge transfer in training processes: towards an integrative evaluation model. *The IUP Journal of Knowledge Management*, 17(1), 7-40.

- Pershing, J. A. (2006). Human performance technology fundamentals. *Handbook of human performance technology*, 5-34.
- Pepitone, J. S. (2002). A case for humaneering: People are not machines, that's why applying engineered work design to knowledge jobs can be a mistake. *IIE Solutions*, 34(5), 39-45.
- Phillips, P., Phillips, J. J., Stone, R., & Burkett, H. (2006). *The ROI fieldbook*. Routledge.
- Ramírez, Y. W., & Nembhard, D. A. (2004). Measuring knowledge worker productivity. *Journal of Intellectual Capital*, 5(4), 602-628. 10.1108/14691930410567040
- Richey, R. C., Klein, J. D, & Tracey, M. W. (2011). *The instructional design knowledge base*. Routledge.
- Rogers, E. M. (2003) *Diffusions of innovation* (5th ed.). Free Press.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17(2), 151-161. DOI: 10.1016/S0886-1633(96)90020-3
- Scriven, M. (1967). *The methodology of evaluation. Perspectives of Curriculum Evaluation*. AERA. Monograph 1.
- Skyttner, L. (2001). *General systems theory: Ideas & applications*. World Scientific.
- Stake, R.E. (2003). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 151–173). Sage Publications.
- Stufflebeam, D. L. (1971, February 24). The relevance of the CIPP evaluation model for educational accountability [Paper presentation]. The Annual Meeting of the American Association of School Administrators, Atlantic City, NJ, United States.
- Tamim, S. R. (2020). Analyzing the complexities of online education systems: A systems thinking perspective. *TechTrends*, 64(5), 740-750. <https://doi.org/10.1007/s11528-020-00538-9>

- Taylor, F. (1911). *The principles of scientific management*. Harper & Brothers.
- Tessmer, M., & Richey, R. C. (1997). The role of context in learning and instructional design. *Educational Technology Research and Development, 45*(2), 85-115. DOI: 10.1007/BF02299526
- Thomas, N., Raghunathan, T. E., Schenker, N., Katzoff, M. J., & Johnson, C. L. (2006). An evaluation of matrix sampling methods using data from the national health and nutrition examination survey. *Survey Methodology, 32*(2), 217.
- Thompson, S. A., & Miller, K. L. (2018). Disruptive trends in higher education: Leadership skills for successful leaders. *Journal of Professional Nursing, 34*(2), 92-96. <https://doi.org/10.1016/j.profnurs.2017.11.008>
- Thorndike, R. M., & Thorndike-Christ, T. M. (2010). *Measurement and evaluation in psychology and education*. Pearson.
- Van Tiem, D. M., & Burns, N. C. (2020). Performance improvement: HPT model. In D. Van Tiem and N. Burns (Eds.), *Cases on performance improvement innovation* (pp. 1-9). IGI Global.
- Van Tiem, D., Moseley, J., & Dessinger, J. (2000). *Fundamentals of performance technology: A guide to improving people, process and performance*. International Society for Performance Improvement.
- Van Tiem, D., Moseley, J., & Dessinger, J. (2012). *Fundamentals of performance improvement: Optimizing results through people, process, and organizations*. John Wiley & Sons.
- Von Bertalanffy, L. (1972). The history and status of general systems. *Academy of Management Journal, 15*(4), 31-38. DOI: 10.5465/255139

Wakefield, A., Spilsbury, K., Atkin, K., McKenna, H., Borglin, G., & Stuttard, L. (2009).

Assistant or substitute: exploring the fit between national policy vision and local practice realities of assistant practitioner job descriptions. *Health Policy*, *90*(2-3), 286-295. DOI:

<https://doi.org/10.1016/j.healthpol.2008.10.011>

Waterman Jr, R. H., Peters, T. J., & Phillips, J. R. (1980). Structure is not organization. *Business Horizons*, *23*(3), 14-26.

Williams, D. D., South, J. B., Yanchar, S. C., Wilson, B. G., & Allen, S. (2011). How do

instructional designers evaluate? A qualitative study of evaluation in

practice. *Educational Technology Research and Development*, *59*(6), 885-907. DOI:

10.1007/s11423-011-9211-

Yazdi, P. G., Azizi, A., & Hashemipour, M. (2019). A hybrid methodology for validation of

optimization solutions effects on manufacturing sustainability with time study and

simulation approach for SMEs. *Sustainability*, *11*(5), 1454-1480. DOI:

10.3390/su11051454

Zinoveff (2008). Review and analysis of training impact evaluation methods, and proposed

measures to support a United Nations system fellowships evaluation framework. WHO's

Department of Human Resources for Health.

VITA

Darden College of Education & Professional Studies

4301 Hampton Boulevard

Norfolk, VA 23529

Chad McLane was born in Downers Grove, Illinois, on November 15, 1980. They graduated from Deer River High School in Deer River, Minnesota, in 1998 and earned a Bachelor of Arts in English with a concentration in literary studies from Brigham Young University-Idaho in 2004; a Master of Arts degree in English from Marquette University in 2007; and a Doctor of Philosophy in Education with a concentration in Instructional Design and Technology from Old Dominion University in 2022. They taught composition, literature, freshman orientation, and study skills courses at Marquette University, Brigham Young University-Idaho, North Iowa Area Community College, and Northeastern State University. Chad was a founding co-director of the NIACC Center for Excellence in Teaching & Learning and served as the director of the NSU Center for Teaching and Learning before returning to Brigham Young University-Idaho as an Online Instructor Manager, where he has worked in department relations, instructor training, instructor hiring, and instructor management.