

ANALYZING THE IMPACT OF OKR GOAL SETTING WITH TEACHERS ON  
STUDENT AND TEACHER PERFORMANCE OUTCOMES

By

Bobby Neal Duke III

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of  
Doctorate in Assessment, Learning, and School Improvement

Middle Tennessee State University

April 2021

Dissertation Committee:

Dr. Kevin Krahenbuhl, Chair

Dr. Lando Carter

Dr. Angela Hooser

Kimberly Osborne

## ACKNOWLEDGEMENTS

Over the past three years I have worked weekend after weekend completing course work and working on my dissertation with the goal of obtaining this doctoral degree. However, to say that none of this would have been possible without those around me is truly an understatement. I want to thank Dr. Krahenbuhl, Dr. Carter, and all the ALSI professors who worked to make every class a great experience. Your guidance and clarity gave all of us in Cohort 6 the confidence we needed to get the job done.

To Neal and Nathan, thank you for being so understanding when I had to write, read, or go to class again for another weekend. Thank you for helping your mom and for letting me reach this goal. I am excited about the weekends of fun to come.

To Leslie, my best friend and biggest cheerleader, I have said from the beginning that half of this degree belongs to you. Thank you for pushing me when I wanted to quit and for making sure I always knew I was supported. You always believed in me and were the rock I needed to get through this. You sacrificed so much so I could do this, and I do not take any of it for granted. There are many things I do not know. However, one thing I know for sure is that I never could have done this without you by my side.

## **ABSTRACT**

School improvement remains one of the most prevalent and often controversial topics in education today, and key to improving schools is understanding the role teachers play as a primary influence on student learning. Consistent and ongoing improvement, however, has not been subjugated to education alone, and looking for improvement strategies outside the field of education can provide valuable insights on new methods of school improvement. In this study, therefore, I investigated the impact of implementing John Doerr's Objectives and Key Results (OKRs) goal setting, improvement model used in business to the K12 school environment.

Following a quasi-experimental design, I conducted a quantitative study that analyzed the impact this improvement model had specifically on teacher performance, student performance in English/Language Arts and Mathematics, and teacher motivation. The literature and research around goal setting and motivation strongly supported the hypothesis that the experimental group would show improvements. However, the data was not supportive of the use of OKRs in any of the research areas.

This research was conducted during the Fall semester of the 2020-2021 school year which coincided with the Coronavirus 2019 (COVID-19) pandemic which greatly altered the schooling environment in many ways. Due to this, the research was impacted by confounding variables that would not have normally been present and results should be viewed through this lens. This research study still provides an insight into the value of goal setting on motivation and performance and provides implications for practice, policy, and future research in this area.

## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	x
CHAPTER I: INTRODUCTION.....	1
Research Problem Statement .....	5
Theoretical Framework .....	6
Purpose Statement and Research Questions .....	7
Summary .....	7
CHAPTER II: REVIEW OF LITERATURE .....	9
Introduction .....	9
The Birth of the OKR System .....	9
Peter Drucker and The Origins of Management by Objectives.....	10
The Creation of iMBOs.....	12
From iMBOs to OKRs.....	14
SMART Goals.....	15
The Evolution of SMART Goals.....	16
SMART Goal Characteristics and Their Alignment with OKRs.....	17
Specific Goals.....	17
Measurable Goals.....	18
Attainable Goals.....	19
Results-Oriented Goals.....	20
Time Bound Goals.....	22
Collaboration and Commitment.....	23

Potential SMART Goal Pitfalls.....	25
Impact of SMART Goals.....	26
The Evolution of Teacher Evaluations.....	27
The Purpose of Evaluations.....	29
Problems in Implementation of Evaluation Systems.....	32
Elements of Effective Evaluations.....	34
Connecting Teacher Evaluations and OKRs.....	35
Goal Setting Theory, OKRs, and Motivation.....	36
Self-Determination Theory.....	38
Autonomous vs. Controlled Motivation.....	39
Competency (Self-Efficacy).....	40
Autonomy.....	42
Relatedness.....	44
The Culture for OKRs to Succeed.....	45
A Foundation of Trust.....	46
Trust and Empowerment.....	48
Psychological Safety.....	49
Expertise.....	52
Deliberate Practice.....	54
Characteristics of Deliberate Practice.....	55
Deliberate Practice and Motivation.....	57
Deliberate Practice and Teaching.....	59

CHAPTER III: METHODOLOGY.....	62
Introduction.....	62
Research Setting.....	63
Methodology.....	65
Rationale.....	69
Participants and Data Sources.....	71
Case English Language Arts Assessment.....	72
i-Ready Diagnostic Math Benchmark Assessment.....	73
Tennessee Educator Assessment Model (TEAM).....	74
Teacher Motivation Survey.....	77
Data Collection Procedures.....	79
Data Analysis Procedures.....	82
Limitations and Delimitations.....	86
Summary.....	87
CHAPTER IV: FINDINGS.....	90
Introduction.....	90
Research Questions.....	90
School and Research Group Profiles.....	91
Research Question 1.....	95
Research Question 2.....	102
Research Question 3.....	104
Research Question 4.....	107

Conclusion.....	119
CHAPTER V: CONCLUSIONS.....	121
Introduction.....	121
Summary of Results.....	123
Discussion.....	125
Teacher Improvement.....	126
Student Performance.....	129
Teacher Motivation.....	132
Environment.....	133
Reflection.....	134
Limitations.....	124
Implications.....	125
Implications for Practice.....	139
Implications for Policy.....	142
Implications for Future Research.....	143
Conclusion.....	144
REFERENCES.....	147

## LIST OF TABLES

1. A Comparison of OKRs and Smart Goals.....	23
2. Percentage of Black, Hispanic, and Native American (BHN) Students.....	64
3. Percentage of Economically Disadvantaged Students.....	64
4. Overview of Steps in the Research Procedure.....	68
5. TEAM Evaluation Domains and Performance Indicators.....	76
6. Independent and Dependent Variables by Research Questions.....	79
7. Numerical Recoding of Motivation Survey.....	85
8. Research Question and Data Summary Table.....	89
9. Percentage of Historically Underserved Students.....	92
10. Characteristics of the Teacher Population.....	93
11. Characteristics of the Teachers in the Experimental Group.....	93
12. Grade Make-Up of Experimental Group.....	94
13. Characteristics of the Teachers in the Control Group.....	94
14. Grade Make-Up of Control Group.....	94
15. Descriptive Statistics for the Population’s Instruction Domain Score.....	96
16. Descriptive Statistics for the Control and Experimental Groups Instruction Domain Scores.....	97
17. Independent Samples t-test for Instructional Domain Rubric.....	98
18. Descriptive Statistics for the Population’s Identified Performance Indicator Focus Area Score.....	99
19. Descriptive Statistics for the Control and Experimental Groups identified Performance Indicator Focus Area Score.....	100



20. Independent Samples t-test for Identified Performance Indicator Focus Area Score.....	101
21. Descriptive Statistics for the Population’s Total T1 and Total T2 Case Scores..	102
22. Descriptive Statistics for the Control and Experimental Groups Total T1 and T2 Case Scores.....	103
23. Independent Samples t-test for Case Assessment Total Score +/-.....	104
24. Descriptive Statistics for the Population’s Total T1 and T2 i-Ready Scores....	105
25. Descriptive Statistics for the Control and Experimental Groups Total T1 and T2 i- Ready Scores.....	105
26. Independent Samples t-test for i-Ready Assessment Total Score +/-.....	106
27. Motivation Survey Questions.....	108
28. Numerical Recoding of Motivation Survey.....	109
29. Descriptive Statistics for the Population’s Motivation Survey Results.....	109
30. Descriptive Statistics for the Control and Experimental Groups Motivation Survey Results.....	111
31. Independent Samples t-test for Motivation Survey Results Question 1.....	112
32. Independent Samples t-test for Motivation Survey Results Question 2.....	113
33. Independent Samples t-test for Motivation Survey Results Question 3.....	114
34. Independent Samples t-test for Motivation Survey Results Question 4.....	115
35. Independent Samples t-test for Motivation Survey Results Question 5.....	116
36. Independent Samples t-test for Motivation Survey Results Question 6.....	117
37. Independent Samples t-test for Motivation Survey Results Question 7.....	118
38. Independent Samples t-test for Motivation Survey Results Question 8.....	119

## LIST OF FIGURES

1. Nonequivalent Control Group Design.....	67
2. Logic Model.....	70
3. TEAM Required Evaluations.....	77
4. Logic Model.....	123
5. OKR Filtering Process of Alignment.....	142

## **CHAPTER 1: INTRODUCTION**

School improvement remains one of the most prevalent and often controversial topics in education today, and there appears to be a clear consensus that changes are needed in our educational system. There is not a clear consensus, however on which improvements are most critical in this journey towards reaching the ultimate goal of improved student learning and producing better prepared graduates. Throughout the country and beyond, numerous education reformers continue to look for and promote various methods and processes aimed at finding the cure to our educational woes. Despite their searching, we have yet to identify and implement clear reform practices that unite policy makers, legislatures, parents, and the teachers. This has led to the wide-ranging implementation of various initiatives that included multiple policy and practical changes. These reform efforts have “spread rapidly over the education landscape, despite any knowledge as to how (or even whether it is possible) to effect improvements envisioned by reform advocates” (Bryk, Gomez, Grunow, & LeMahieu, 2017, p. 5).

The journey towards school improvement can be traced back through multiple policy changes at the federal level which has set the context for where we find ourselves today. In 1965, the passage of the Elementary and Secondary Education Act (ESEA) set out to improve schools by providing additional federal money designed specifically to meet the needs of disadvantaged students. As Tucker, (2014) emphasizes, the original ESEA law seemingly pointed out that if students were not learning as they should, then the issue was found in the background which students came to school with and additional funds could help compensate for these challenges. However, as the cost to operate

schools and per pupil spending increased, scores among the nation's students made little growth (Tucker, 2014).

In 2001, the No Child Left Behind Act (NCLB) ushered in and promoted the practice of school accountability. NCLB focused reform efforts on improving the school so that they could reach set benchmarks of student proficiency with the ultimate goal of 100% of students performing on grade level. In 2009, the focus of school improvement narrowed from ensuring school success to ensuring teacher success. With the passage of President Obama's Race to The Top program, states were encouraged to implement reforms that tied directly to the teacher, such as the teacher evaluation process, as a lever for improvement (Tucker, 2014). These pieces of federal legislation have encouraged the shifting of reform efforts from addressing students, to the school itself, to ultimately the teacher in the classroom.

The focus on the classroom teacher and their importance in student learning is not just an idea reserved for policy makers. The direct impact of the teacher on student success continues to be supported by research, and it often points to them as the key to improvement. Hattie (2012) clearly states that "teachers are among the most powerful influences in learning" (p. 22). In addition, Darling-Hammond and Goldberg (2001) emphasizes that "the single most important determinant of success for students is the knowledge and skills of that child's teacher" (p. 689). With this in mind, it is clear that school improvement must include a focus on teacher improvement in some capacity. In addition, school leaders must be equipped to lead these efforts in a way that ensures the improvements benefit not only the goals of the individual teacher but the goals of the school as a whole.

Bryk et al. (2017) states that “if we continue to seek improvement in the ways we have always done, we are likely to continue to get what we have always gotten” (p. 6). With this in mind, it is important to acknowledge that improvement practices and the goal of producing better results is found in all occupations and is not relegated just to education. Therefore, looking outside of traditional educational reform models may provide an avenue for school improvement. The business sector, specifically, has a rich history of literature designed to provide managers a guide to producing better outcomes. One of the original and most prominent voices in the business improvement effort was that of Peter Drucker who has been referred to as the management theory’s greatest thinker and the father of modern management (Bell, 1999). His work on objective-based management became a seminal influence on the field (Zahra, 2003) and shifted the conversation from the activities employees were involved in to understanding the purposes of the activity itself and the setting of goals to be achieved (Bell, 1999).

Today’s managers and leaders, however, are working increasingly with employees whose jobs require them to work in consistently changing and unknown contexts rather than in the highly predictable contexts of the past. Therefore, in considering Drucker’s work and his emphasis on meeting objectives as a road map to continuously improve, one must also consider the key role that employee motivation plays in the process. As Pink (2009) emphasizes, there is a difference between the motivation needed for algorithmic tasks where there is a single pathway to a clear conclusion and heuristic tasks where there is not clear algorithm and devising novel solutions is needed. Whereas algorithmic tasks may be motivated by extrinsic rewards, heuristic tasks require a more intrinsically motivated employee (Pink, 2009). Therefore,

leaders must ensure they are coupling the right form of motivation with the task and area of improvement they are seeking.

One specific management methodology that has become increasingly popular in the continuous improvements of businesses and has the potential to aid in the fields like education is Doerr's (2018) process of Objectives and Key Results (OKRs). OKRs are built off the work of Peter Drucker and Andy Grove who both emphasized the importance of managing by objectives.

Although goal setting with teachers is not a novel practice in schools today, it has typically been viewed as a team activity focused on raising student achievement levels instead of an independent challenge based on the teacher's needs.

The most popular form of goal setting in schools is conducted through the professional learning community (PLC) process which emphasizes the creation of yearly SMART (strategic, measurable, attainable, results oriented, time bound) goals for like teams throughout the school (DuFour, DuFour, Eaker, Many, & Mattos, 2016). Although there are similarities between SMART goals and OKRs, the OKR system provides a more complete framework for employees and their leaders. OKRs emphasize the importance of on-going feedback, transparency, and clear key results to measure the progress of the work. In addition, the self-selection of OKRs is critical in supporting employee autonomy and increasing employee motivation.

Doerr (2018) clearly articulates the benefits that companies have seen by implementing this process and how they use OKRs as a primary driver for improvement. This raises the question that if this clear framework can aid in the private business sector, could it also work in the public-school environment? Implementing OKRs in a school

setting could help bridge the personal gap that teachers may feel if they are only using the PLC SMART goal process. Whereas PLC goals are typically team developed and focused solely on raising achievement scores, ORKs provide the opportunity for teachers to set individualized goals that can target not just raising achievement levels but the specific strategies, methods, or teaching practices that will aid in the raising of achievement. For example, a typical SMART goal may be for 65% of students to show proficiency on a math benchmark or summative assessment. However, an OKR could more directly tie to the practices that will help achieve this. This could include a teacher seeing the need to increase the amount of student discourse within the class and therefore create an OKR around this strategy. This goal setting process can work to make improvement efforts intentionally personal to the educators and their individualized improvement efforts as they work to move both their students and their teaching craft forward.

### **Research Problem Statement**

It is clear that teacher improvement must be a substantial part of the school improvement process. However, with so many reform initiatives at play, there is the danger that teachers may be unclear on what improvements they should be striving for and what success in the area looks like. If improving teacher practice is the key to improving schools and student learning, it is imperative they are provided a clear roadmap on how to improve.

## **Theoretical Framework**

Embedded within the OKR process are characteristics that are clearly associated with two research-based theories that will help support the implementation of OKRs as a strong possibility for improving student achievement and teacher performance.

First, Locke and Latham's (1990) goal setting theory has been used to illustrate the connection between the setting of goals and an individual's task performance. Specifically, the theory emphasizes that value of individuals setting specific goals that are high but attainable and the impact it has on a higher performance.

A second theory that supports this implementation is Ryan and Deci's (2000) Self-Determination Theory (SDT) which connects OKRs to increasing the intrinsic motivation of employees as they work to reach goals. Key characteristics of the OKR process include the joint development of goals between employee and leader, on-going feedback, and the ability to adjust goals as needed. In addition, these practices increase the capacity to create cultures built on mutual trust and psychological safety. These characteristics support an individual's autonomy, competency, and relatedness which SDT emphasizes is not only needed to build intrinsic motivation but are also basic human needs.

As applied to this study, these two theories support that by setting OKRs teachers would not only be more aware of improvement areas and the actions needed to accomplish the improvements, but they would also be more intrinsically motivated to reach their stated goals.



## **Purpose Statement and Research Questions**

Doerr's (2018) book *Measure What Matters* provides a context around the implementation of the OKR system in various businesses, and it details the success these businesses have seen because of the implementation. The purpose of this research is to collect quantitative data to evaluate and inform on the impact of applying the OKR process as a method for school improvement. In this study, I will investigate improvements as measured by both teacher improvement, student academic improvement, and teacher motivation improvement to answer these questions:

1. What impact does the creating and monitoring of individual, OKR goals have on teacher performance as measured by teacher evaluation data?
2. What impact does the creation and monitoring of individual, OKR goals have on student learning in the English/language arts content area as measured by student performance data?
3. What impact does the creation and monitoring of individual, OKR goals have on student learning in the mathematics content area as measured by student performance data?
4. What impact does the creation and monitoring of individual, OKR goals have on teacher motivation as measured by perception survey data?

## **Summary**

The following chapters outline the background information important to the research as well as outcomes and implications for the profession. In chapter 2, I provided an historical overview of the OKR system and lay out the similarities and differences between OKRS and the PLC driven SMART goals which are widely used in schools. In

addition, I explore whether the teacher evaluation system, which is presently used for holding teachers accountable and identifying the support they need to improve, is meeting its intended goals. I further provide an analysis of motivational theories and cultural elements that support the development of intrinsic motivation and detail its value in the continuous improvement cycle. Finally, I connect OKRs with the theories of deliberate practice as a model for improvement. Chapter 3 provides a clear, step-by-step process of the methodology used in this quantitative study. In Chapter 4, I will reveal the results of the data analysis for the teachers who participated in the research compared to those who did not. Finally, in Chapter 5, I will discuss the implications for practice and policy moving forward as well as recommendations for future research.

Ultimately, this research establishes a clear protocol and method for educators and leaders to use in their continued efforts to improve the teaching and learning occurring in schools. The process honors the complex nature of the teaching profession and provides an avenue to set clear objectives, obtain feedback, and clearly measure success.

## **CHAPTER II: REVIEW OF LITERATURE**

### **Introduction**

Dylan Wiliam (2018) reminds us that the “job of teaching is so difficult, so complex, that one lifetime is not enough to master it” (p. 29). Therefore, it is incumbent upon teachers, administrators, and all educators to identify and put into action proven techniques that are shown to improve teacher practice so that the ultimate goal of positively impacting student achievement can be attained. This literature review establishes a foundational platform for the application of a business protocol for operation improvement called Objective and Key Results (OKR) to the educational system. Throughout the chapter, peer-reviewed empirical research and theories that both connect and support the OKR process are examined. Particular attention is paid to current prominent practices around goal setting in education, teacher evaluation practices, increasing motivation, building a school culture based on trust and psychological safety, and increasing the opportunities for teachers to improve their abilities through the use of deliberate practice.

### **The Birth of the OKR System**

John Doerr (2018) states that, “Ideas are easy. Execution is everything” (p. 6). This mantra lays the foundation for his system of management by Objectives and Key Results or OKRs. Throughout his book *Measure What Matters* (2018), Doerr lays out how the OKR system, that is built on the creation of bottom-up, measurable, and public key results, have helped organizations like Google, YouTube, Lumeris, and The Gates Foundation stay successful and profitable year after year. Although Doerr (2018) does

state that he assembled the acronym OKR, he clearly points to others before him who provided the guidance on which this work is built. While serving as an engineer at Intel in the 1970s, he worked under Andy Grove who he claims was “the greatest manager of his or any era” (Doerr, 2018, p. 6). Doerr (2018) points to Grove as the father of OKRs, but to see the true inception of the process one must go back further to another great thinker in business management, Peter Drucker.

### **Peter Drucker and The Origins of Management by Objectives**

Born in 1909 in Vienna and educated in England before moving to the United States in 1937, Drucker was asserted to be one of the most influential thinkers on management of his era (Zahra, 2003). Bell (1999) points out that Drucker began studying management in the 1940s, and throughout his career, which lasted over six decades, he became regarded as the father of modern management (p. 35). Drucker’s management expertise is built around the concept of management by objectives (MBOs), and although he never claimed the distinction, he is credited with inventing the term and was the first to publish the concept (Greenwood, 1981). In 1950, Drucker was hired as a consultant for the reorganization of General Electric (GE), and it was during this time through his work with Harold Smiddy at GE that he first implemented the practice (Greenwood, 1981). This work led to the publishing of his book *Practice of Management* in 1954 which popularized the concept (Vohra & Mukul, 2009).

Management by objectives is considered Drucker’s greatest contribution to business strategy, and it shifted the conversation from process to goals and from the completion of activities to the purpose behind the activities (Bell, 1999). Ford, McLaughlin, and Nixdorf (1980) provide the following definition of MBO:

A managerial system that seeks to involve the individual employee through an interactive process of goal setting, participation and feedback. The manager and subordinate collaborate in developing specific measurable goals, which are to be accomplished in a specific period of time. In this collaboration, the managers seek to merge the talents and job responsibilities of the subordinate with the overall needs of the organization. (p. 88)

Drucker emphasized that objectives should be the starting point for planning, and the real role of managers was in setting the objectives and in deciding what the business is, what it should be, and what it could be (Greenwood, 1981). He articulated that objectives were not given and known to all, and Greenwood (1981) states, “Peter Drucker put objectives into center stage and made them the core of the structure of a discipline of managing” (p. 230).

The role of the employee in setting of goals was a key piece to the MBO system. In Drucker’s system, subordinates played a major role in the creation of their personal objectives instead of it coming from supervisors (Greenwood, 1981). Vohra and Mukul (2009) point out this assisted in combating what Drucker called the activity trap. In addition, this inclusion underscored his emphasis that managers have to focus on “performance, people, and motivation (Vohra & Mukul, 2009, p. 2). According to Doerr (2018), Drucker understood a “basic truth of human nature: When people help choose a course of action, they are more likely to see it through” (p. 25). This model with its humanistic approach to results-driven management went in the opposite direction of former authoritarian management styles of the past (Doerr, 2018).

Vohra & Mukul (2009) detail how in 1960, Drucker coined the term “knowledge worker” to describe a growing population of employees who were paid to work with brains instead of their hands. They explain that Drucker understood this group of workers needed a new type of management that emphasized the motivation of the employee and the importance of them being willing to take responsibility for the whole. This growing cadre of workers created an environment where the role of the manager shifted to the creation of contexts where employees take responsibility for their work (Vohra & Mukul, 2009). In addition to the inclusion of employees on company goals, Drucker’s MBOs focused on other key aspects including a balance of both short and long-term planning that was informed by data and having regular conversations with colleagues about their progress towards goals (Doerr, 2018). These practices “became Andy Grove’s foundation and the genesis of what we now call OKR” (Doerr, 2018, p. 25).

### **The Creation of iMBOs**

Born in Hungary in 1936, Andy Grove began his studies at the University of Budapest. However, he left his family and his homeland for the United States when the Hungarian Revolution opened his country’s borders in 1956 (Geppert, 2000). Geppert (2000) discusses his transition to the United States which included learning English from other Hungarian immigrants, overcoming hearing loss, and navigating the unfamiliar teaching methods that differed from those in Hungary’s universities. However, by the late 1960’s, Grove had earned a Ph.D. in chemical engineering and had started working at Fairchild Semiconductor (Tedlow, 2005). It was here that he met co-workers Robert Noyce and Gordon Moore, and he would follow them to start Intel in 1968.

At Intel, Grove was put in charge of operations, and this new position also gave Grove the new experience of having to lead other people (Tedlow, 2005). Tedlow (2005) says that Grove took his initial management lesson and created his job description from a Time magazine article about movie directors titled “Vision to Inspire.” The article detailed the importance of mastering complexity and fusing individual elements into a whole. Grove’s managerial talent was discovered early, and he was said to have strong reasoning abilities that allowed him to reduce large sets of facts to logical information (Geppert, 2000).

Grove rose in the Intel ranks to executive vice president and eventually CEO. Doerr (2018) points out that Grove made a leap from Drucker’s work by applying the manufacturing principals to softer professionals made up of people who were paid to think. He focused on creating environments that emphasized output and specifically on separating activity from output. Inspired by Drucker, Grove created his goal-setting system and named it “‘iMBOs,’ for Intel Management by Objectives” (Doerr, 2018, p. 26). Tedlow (2005) emphasizes that Grove’s success is tied to the fact the he approached management as a discipline, and that he was, “America’s greatest student and teacher of business” (p. 116).

One of the differences between Drucker’s MBOs and Grove’s iMBOSs was the focus Grove put on key results. As Doerr (2018) states, “Grove rarely mentioned objectives without tying them to ‘key results,’ a term [Grove] seems to have coined himself” (p. 26). Grove defined objectives as the direction that lay out where you are going. Key results, however, are separate and measure the steps toward the objectives achieved (Doerr, 2018). Other differences between this new and old system include

moving from annual objectives to quarterly or monthly, from having private objectives to public and transparent ones, increasing the emphasis on bottom-up or sideways creation or goals, and divorcing them from compensation (Doerr, 2018). Grove's emphasis on key results and outputs is what set Intel apart from other companies who tended to value experience regardless of whether or not that led to results. This attitude was captured in their slogan "Intel delivers" and reiterated the fact that knowledge was second to execution (Doerr, 2018).

### **From iMBOs to OKRs**

John Doerr worked under Andy Grove at Intel and became a student to his iMBO process. In his book, *Measure What Matters* (2018), Doerr chronicles his process of taking the teachings of Grove and using them with companies from around the world under the acronym he coined, OKRs. Doerr (2018) defines objectives as "simply WHAT is to be achieved" and key results as the process to "benchmark and monitor HOW we get to the objective" (p. 7). Where objectives are significant, concrete and ideally inspirational, key results are specific, time bound, measurable, and verifiable. With key results there is a clear distinction in that you either met it or you did not. "There is no gray area, no room for doubt" (p. 7) with whether or not key results were met. Together, objectives and key results provide "A management methodology that helps to ensure that the company focuses efforts on the same important issues throughout the organization" (p. 7) Doerr (2018) points out that "while conceptually simple, Andy Grove's regimen demands rigor, commitment, clear thinking, and intentional communication (p. 13).

Doerr states that ultimately, his goal is to see the OKR system not just as a process but as a launchpad and to "see Grove's brainchild transform every walk of life"



(p. 245). One of these areas specifically mentioned by Doerr is in the success of school systems. Currently, however, a different goal setting process has dominated much of the educational world as professional learning communities (PLCs) have gained popularity and prominence.

### **Smart Goals**

Richard DuFour and Becky DuFour (2008), architects of the Professional Learning Community (PLC) process, emphasize that educators can make one of two basic assumptions about teaching and learning. The first is that student learning is based on the effort and ability of the students and therefore teachers and schools have no impact on learning. The second assumption lays opposite of the first and states that teachers and schools can have significant impacts on student learning. This second assumption emphasizes that as teachers work more effectively, they can help more students learn at high levels. The PLC process is driven by a key idea that educators must work together collaboratively and take a collective responsibility for student learning. It emphasizes a belief in the second assumption mentioned above (DuFour et al., 2016). This key idea along with the emphasis PLCs place on being results oriented has made the creation of SMART goals a central part of the PLC process.

DuFour et al. (2016) define SMART goals as goals that are “Strategic and specific, Measurable, Attainable, Results oriented, [and] Timebound” (p. 90). They emphasize that SMART goals are not only an important part of the PLC process, but they are “the best way to help people throughout a school district to truly focus on results” (DuFour et al., 2016, p. 89). The PLC movement has increased rapidly since 2000 and with its rise has come widespread agreement on the power of PLCs as an effective school

improvement model (Eaker & Keating, 2012). This means that more schools around the country are implementing SMART goals as part of their ongoing instructional practices. However, the use of SMART goals is not only isolated to schools and has its origins outside of the school walls.

### **The Evolution of SMART Goals**

The origins of the term SMART goals differ based on various literature, but researchers often attribute Edwin Locke's (1968) work around motivation and incentives in the business world as providing the key components of the SMART goal process (Brown, Leonard, & Arthur-Kelly, 2016). Locke's theories in 1968 corresponded with the opening year of Intel and would also prove as a strong influence on Andy Grove (Doerr, 2018). This important connection demonstrates that both the OKR and SMART goal process have the similar foundation of Locke's (1968) work.

Education highlighted the importance of goal setting prior the emergence of the PLC process. However, works by McGreal (1980) focused more on the importance of goal setting in a supervisory role to help clarify expectations for teachers. Ultimately, it has been the emergence of the PLC process that has brought SMART goals to regular use in education.

Camp (2017) states that the term "goal" can have different meanings depending on the context and states that under the broad definition goals can "encompass the grandiose and the mundane, the deliberate and the subconscious" (p. 61). However, based on over four decades of goal setting theory, an established definition for a goal is "the object or aim of an action" (Locke & Latham, 2002, p. 705). Even with this common definition, MacLeod (2012) states that the use of the term "goal" when referring to

SMART goals can actually be a misnomer. As Doerr (2018) does with OKRs, MacLeod (2012) ties the origins of SMART goals to the work of Peter Drucker's business model of management by objectives (MBO), and therefore states that the way the term goals is used where referring to SMART goals actually refers more to objectives. MacLeod (2012) points out that where traditional goals are broad and general, objectives are narrower in scope and specific. Furthermore, he points out that traditional goals tend to be abstract and hard to validate whereas objectives are easy to validate and concrete. This distinction is important as DuFour et al. (2016) emphasize the importance of clarity when creating the SMART goals that should be used to drive the work of schools. The characteristics of objectives as shared by MacLeod (2012) are more in line with the DuFour et al. (2016) definition of SMART goals as used in the PLC process as they are specific, measurable, attainable, results oriented and time bound.

### **SMART Goal Characteristics and their alignment with OKRs**

#### *Specific Goals*

DuFour et al. (2016) points out that when a team writes an effective goal it will help answer the question, "How will we know if our strategies are resulting in gains in student learning?" (p. 93). He points out that the first step in the process is ensuring specificity on what you are trying to achieve. Making objectives specific helps distinguish efforts from results (MacLeod, 2012). Locke and Latham (2002) also emphasize the importance of specificity in goal setting. Their research indicates that goals need an external referent otherwise success is subjective and defined individually. They emphasize, when people are asked to do their best, they do not.

OKR's similarly share a focus on specificity in the writing of key results. Doerr (2018) emphasizes that disciplined thinking is the key to effective goals, and the process begins with the question, "What is most important for the next three (or six, or twelve) months?" (p. 47). Specificity in goal setting is key because it reduces the variance in possible performance and reduces ambiguity about what is trying to be obtained (Locke & Latham, 2002). In addition, specificity provides clarity around the goal which Camp (2017) points to as a way to increase commitment to the goal.

### *Measurable Goals*

Creating goals that are measurable helps to ensure teams reach the core component of the PLC process of mutually accountability (DuFour et al., 2016). MacLeod (2012) echoes the connection between measurement and accountability, and stresses that measurable criteria help ensure goals are not subject to interpretation and eliminates the possibility of disagreements. In addition, measurement criteria provide a way for teachers to track their progress towards the goal. Teachers are familiar with engaging in this process with students through progress monitoring and formative assessments practices (Brown & Arthur-Kelly, 2016), and Camp (2017) points to findings that an individual's awareness of their progress towards a task is an essential component for goals to be effective.

Similarly, OKRs also put a strong focus on the ability to measure the key results that have been chosen. The ability to clearly indicate whether or not a key result has been achieved is one of the most important characteristics of an effective OKR (Doerr, 2018). This is why Doerr (2018) points out that key results almost always include a number to serve as a gauge of progress. As discussed previously, there should be little

gray area or room for doubt when considering if the key result has been met. With OKRS, verifying progress is essential (Doerr, 2018).

Chamberlin (2011) summarizes the importance of this component of the SMART goal process by reminding us that, “If you can’t measure it, you can’t manage it, so you won’t know when you are done” (p. 23). SMART goals are not missions, and people need to know when the goal has been accomplished (Chamberlin, 2011).

### *Attainable Goals*

Depending on your source, the A in SMART goals can stand for either achievable or attainable. In the context of PLCs, attainable is most often used. As mentioned previously, mutual accountability is a key component of the PLC process, therefore the attainability of the goal is central to helping ensure the fairness of the accountability. Goals must be aligned to the available time, talent, and resources to avoid frustration (MacLeod, 2012). The intent of attainable goals is to build momentum and help secure short-term wins which is key to sustaining the effort towards accomplishing goals (DuFour et al., 2016).

Even though attainability is key, Locke and Latham (2002) underscore the value of setting goals that are more difficult to accomplish versus those that are easier to accomplish. Their work points out that when more difficult goals are produced there is a higher level of effort and performance in working to meet that goal. As they emphasize, goals have an “energizing function” (p. 706) which is why high goals lead to greater effort. In determining how difficult to make a goal, DuFour et al. (2016) emphasize that, “Teams should feel reasonably confident they have the capacity to achieve their goals. They should be able to say, ‘If we seek and implement best practices, we have reason to

believe we will achieve our goal” (p. 92). In addition, DuFour et al. (2016) underscores the connection between attainability and being results oriented, the R in SMART.

For OKRs, the accountability aspect of goal setting is done largely through the emphasis on keeping a limited number, three to five, key results for each objective (Doerr, 2018). When too many key results are set, Doerr (2018) states there is danger that one can lose focus or obscure the progress being made. This would directly impact the important elements of specificity and measurability. Doerr (2018) also encourages the writing of short term OKRs, usually quarterly. He states, “it’s the shorter-term goals that drive the actual work. They keep annual plans honest – and executed (Doerr, 2018, p. 51).

The concentration on attainable does not, however, indicate that goals should be achieved with little effort. Both PLC driven SMART Goals (DuFour et al., 2016) and OKRs (Doerr, 2018) emphasize that goals should push those who set them, and it is the leader’s responsibility to help teams balance between more easily attainable goals and stretch goals that help “stimulate creativity and innovation” (DuFour et al., 2016). Research (Locke & Latham, 2002) indicates that high or more difficult goals do lead to higher levels of effort and performance. However, leaders must ensure their employees understand not only the importance of the outcome but also the belief that outcome is attainable (Doerr, 2018).

### *Results-Oriented Goals*

DuFour et al. (2016) clearly state that the R in SMART stands for result-oriented goals, and they specifically warn against goals that confuse developing a plan with taking meaningful action to ensure something actually occurs. Results-oriented goals emphasize

that the purpose is to focus on improved results for students instead of simply implementing activities, and that goals should focus directly on the evidence of student learning instead of goals focused on what teachers will learn. (DuFour et al. 2016). This supports what O’Neill (2000) states is the real power in setting smart goals, “learning whether teachers make a difference in student learning” (p. 48). She continues by stating the SMART goals help educators understand how effective their instructional programs and processes are and in quantifying direct results which provides educators with better feedback on their progress.

Although SMART goals are used both in education and in business, defining the R is one place where there can be a sharp divide. There have been other emphasized meanings for the R in SMART including “realistic” (Brown et al, 2016), “reasonable, or relevant” (Chamberlin, 2011). In his writing relating to SMART goals in business, Chamberlin (2011) argues against the use of result oriented. He claims that, by definition, all goals are result-oriented, and if your goal is specific, measurable, and attainable then you will achieve it. He instead urges the creation of goals that are relevant to your purpose as a company and the needs of your customers (Chamberlin, 2011). OKRs seemingly mirror Chamberlin’s (2011) understanding of goals in two ways. First, OKR’s emphasis on results is embedded in both the name and the process. Second, OKRs highlight the importance of alignment in writing effective goals. Alignment is about linking the work of individuals to the organization’s vision. When OKRs are aligned it assists in narrowing down what is urgent and needs to be done from the task that may call for one’s attention (Doerr, 2018).

The discrepancy between OKRs and PLC driven SMART goals here may be insignificant as DuFour et al. (2016) do emphasize the importance of goal alignment in linking the teacher, school and district goals. However, a results orientation is not just important to the creations of SMART goals in the PLC process, but to the PLC process itself. “Members of a PLC recognize that all of their efforts must ultimately be assessed on the basis of results rather than intentions” (DuFour et al., 2016, p.12).

#### *Time Bound Goals*

MacLeod (2012) promotes the use of the term time bound over timely which has been used in other SMART goal models. He also emphasizes that time bound indicates the mutual agreement of when the objectives are to be accomplished. DuFour et al. (2016) agrees that to say a goal should be time bound indicates that there should be an agreed upon time frame within which the goal will be achieved. Likewise, OKRs also focus on time bound as a characteristic of effectiveness (Doerr, 2018). When goals or objectives do not include this important factor, there is a lack of discipline related to priorities and time management (MacLeod, 2012). MacLeod (2012) further emphasizes that when goals are not time bound there is a “less than rigorous pursuit of closure” (p. 70). This aligns with Locke and Lathom (2002) research that indicates goals can affect participants’ persistence. Specifically, they state, “Tight deadlines lead to a more rapid work pace than loose deadlines” (p. 707).



Table 1

*A comparison of OKRs and SMART Goals*

<b>PLC SMART Goals</b>	<b>OKRs</b>
<ul style="list-style-type: none"> <li>• Specificity on what is trying to be achieved.</li> <li>• Distinguishes efforts from results</li> </ul>	<ul style="list-style-type: none"> <li>• Specificity found in writing key results</li> <li>• Reduces variance in performance</li> </ul>
<ul style="list-style-type: none"> <li>• Measurable goals help ensure mutual accountability</li> <li>• Ensures goals are not subject to interpretation</li> </ul>	<ul style="list-style-type: none"> <li>• Clearly indicated whether or not a key results have been achieved.</li> <li>• Verifying progress is essential</li> </ul>
<ul style="list-style-type: none"> <li>• Attainability is central to mutual accountability</li> <li>• Alignment of time, talent, and resources</li> <li>• Goals should push those who set them</li> </ul>	<ul style="list-style-type: none"> <li>• Emphasis on limiting the number of key results for objectives</li> <li>• Focusing on 3 – 5 key results helps ensure one does not lose focus</li> <li>• OKRs should push those who set them</li> </ul>
<ul style="list-style-type: none"> <li>• Results-oriented goals focus on improved results as the primary purpose</li> <li>• Results focus on improved results for students and evidence of student learning</li> </ul>	<ul style="list-style-type: none"> <li>• Focus on relevance and alignment of goals with the organization’s vision</li> </ul>
<ul style="list-style-type: none"> <li>• Time-bound goals indicate agreement on a time-frame where the goal should be achieved</li> </ul>	<ul style="list-style-type: none"> <li>• Time-bound goals indicate agreement on a time-frame where the goal should be achieved</li> </ul>

**Collaboration and Commitment**

A key part of SMART goals as laid out in the PLC process is the emphasis on team goals and collaboratively writing the objectives to be accomplished. DuFour et al. (2016) lays out a reciprocal nature between SMART goals and effective PLCs. He emphasizes that PLCs are the optimal setting for the creation of truly meaningful goals

while creating meaningful SMART goals helps to develop powerfully collaborative teams. When created collectively, SMART goals both promote ownership of the goals and create a sense of challenge and meaningful purpose (MacLeod, 2012). When people are allowed to participate in setting goals versus being assigned goals by supervisors, they set higher goals and perform at a higher rate (Locke & Latham, 2012). This employee driven creation of goals found in the PLC process clearly aligns to the importance of employee created OKRs discussed earlier. As Doer (2018) points out, “An optimal OKR system frees contributors to set at least some of their own objectives and most or all of their key results” (Doerr, p. 88). This further confirms that collective participation of all team members towards a mutual goal is a vehicle for moving teacher practice and student learning forward.

Much like the PLC process, OKRs also stress collaboration and teamwork in writing goals. However, OKRs have a different focus. For OKRs, collaboration is ultimately about transparency and ensuring everyone understands what key results are being worked on by others on the team or in the company (Doerr, 2018). According to Doerr (2018), this transparency provides two benefits. First it provides an element of accountability which increases motivation for achieving the goal. Second, it provides an avenue for teams to assist each other and to better align the work of the whole. “When goals are public and visible to all, a ‘team of teams’ can attack trouble spots wherever they surface” (Doerr, 2018, p. 89). This is specifically beneficial as public goal setting is a factor that has been shown to increase commitment to the goal (Camp, 2017), and when people are committed to the goal, the relationship to performance is strongest (Locke & Latham, 2002). Locke and Latham (2002) connect public goal setting to commitment

because it connects the participant's actions to their integrity. MacLeod (2012) emphasizes that although this may take more time and greater interpersonal skills it also results in a much higher probability of success.

### **Potential SMART Goal Pitfalls**

Although SMART goals are widely used in both business and PLCs within schools there are areas of concern that can be associated with the process. Brown and Arthur-Kelly (2016) point to evidence suggesting that many teachers may struggle writing effective goals that meet the SMART criteria and with how to show progress towards the goal. Brown and Arthur-Kelly (2016) go on to say that "writing SMART goals appears to be a complex problem-solving process for many teachers" (p. 629). Once goals are established, Camp (2017) identified time as a potential problem area as well for teachers in this process. Her research showed that participants cited time constraints as a central reason for not completing goals. The research suggests that although they may see goal setting as a positive activity it has a lower priority when weighed against other requirements (Camp, 2017). When discussing SMART goals outside of the educational lens, MacLeod (2012), points to potential problems such as having a fixed focus during what could be changing times and a lack of feedback on progress towards the goal. Leadership can play a pivotal role in helping to mitigate these concerns. Flexibility and real-time feedback are suggested parts of the goal design process (MacLeod, 2012) along with a focus on giving explicit guidance and support at all points in the process and in facilitating strong professional learning, (Brown and Arthur-Kelly, 2016).

Likewise, the OKR system is not without its fair warnings. It is possible for OKRs to be implemented poorly which could lead to “a narrowed focus, unethical behavior, increased risk taking, decreased cooperation, and decreased motivation (Doerr, 2018, p. 9). However, Doerr (2018) emphasizes that although leaders must carefully watch for these potential traps, goals continue to be a necessity for high performance.

### **Impact of Smart Goals**

In addition to the impact goal setting has on clearly articulating and accomplishing objectives, there are other benefits that have been found from goal setting. Camp (2017) found that not only did teachers find the goal setting was a positive investment in time but that it also led to perception that their teaching had improved. This is important as Locke and Latham (2002) emphasized the importance of strong self-efficacy in the continuation of goal setting and of setting higher and more demanding goals. In Camp’s (2017) study, even teachers who did not make significant progress on their goal expressed that the process was valuable as it facilitated purpose-driven reflection and problem solving. SMART goals and OKRs are about clearly defining where the team is going and how they will know when they have arrived, and according to Lencioni (2002), “an unrelenting focus on specific objectives and clearly defined outcomes is a requirement for any team that judges itself on performance” (p. 216).

However, for teachers, the official measurement of their performance lays largely outside of the goal-setting process. Instead, states and districts are implementing specific evaluation systems that have very different characteristics than a reliance on achieving specific and aligned goals.

## **The Evolution of Teacher Evaluations**

Conversations and discussions on improving the educational system in our country often include several key topics that have been consistently used as levers for school improvement for many years. One of these levers has been teacher evaluation systems and their use to drive student achievement and teacher performance. Darling-Hammond (2014) states that she first began studying these systems in the 1980s, and in those early years little evidence was found of effective evaluation systems that improved teaching and learning. However, since her initial research, changes in federal laws and the race for federal funds have put new attention on the process and purpose of accountability and the teacher evaluation system (Darling-Hammond, 2014).

Danielson (2001) traces the focus on teacher quality and therefore teacher evaluations back to the late 1990s which she states was actually the third phase of the modern school reform movement. This phase followed an initial phase focused on the length of the school year and academic courses and a second phase which began emphasizing academic standards and the high-stakes assessments that measured them. This third phase started with the publication of a report from the National Commission on Teaching and America's Future titled *What Matters Most: Teaching for America's Future* in 1996. According to Danielson (2001), this third phase made it clear to everyone what many already understood: "the quality of the teacher matters" (p. 12).

With the passage of the No Child Left Behind Act of 2001 (NCLB) which reauthorized and revised the Elementary and Secondary Education Act of 1965 (ESEA), accountability has become a central part of the American educational system. NCLB's purpose was to provide a "fair, equal, and significant opportunity to obtain a high-quality

education” (Davis, Lampley, & Foley, 2016, p. 45), and the focus for this law was on the school’s ability to obtain proficiency on state achievement measures. As Tucker (2014) points out, this new law was an extremely aggressive move by the federal government in tying federal dollars to accountability measures created by the state to ensure certain academic benchmarks were met by schools by 2014.

Teacher evaluations and teacher evaluation reform took more prominent attention in 2009 when the Obama administration passed the American Reinvestment and Recovery Act (ARRA). Kraft and Gilmour (2017), point out that the passage of the ARRA law in 2009 coincided with a growing national concern that the teacher evaluation system was in large parts broken based on a 2009 report from the New Teacher Project (TNTP) titled *The Widget Effect* (Weisberg, Sexton, Mulhern & Keeling, 2009) which had been recently released. The report highlighted the importance of teacher effectiveness, but also stated there was a “long-standing failure to recognize and respond to variations in the effectiveness of teachers” (Weisberg et al., 2009, p. 32). The authors of this study pointed out that teacher evaluation systems were doing little to assess variations in teacher performance but were instead being used to “assume effectiveness is the same from teacher to teacher (Weisberg et al., 2009, p. 32). ARRA provided 4.35 billion dollars in federal funds to the U.S Department of Education to enact a competitive grant entitled Race to the Top (RTTT) (Davis et al., 2016). This new competitive grant along with state waivers for the NCLB regulations created strong incentives for states to adopt changes to the teacher evaluation systems (Kraft & Gilmour, 2017). RTTT moved states to high-stakes evaluation systems that included multiple rating categories, measures of teacher performance, annual observations with feedback, and the use of

evaluations to inform staffing decisions (Connally & Tooley, 2016). Tucker (2014) emphasizes that the changes implemented from NCLB to RTTT did not just prioritize changes to teacher evaluations but moved the conversation from holding schools accountable to holding individual teachers accountable for student learning.

### **The Purpose of Evaluations**

Discussions around teacher evaluation should begin with an understanding of the primary purposes for the teacher evaluation system. Danielson (2001, 2010), Darling-Hammond (2014), and Connally and Tooley (2016) all point to the two primary purposes: accountability or quality assurance and professional development or teacher support. However, it has been argued that previous evaluation systems have failed at accomplishing either goal (Danielson, 2001).

State laws driven by federal incentives have emphasized the importance of the use of teacher evaluations for quality assurance purposes. This first purpose is largely the only purpose considered by policy makers and legislators who are tasked to ensure that public funds used to educate the community are helping to guarantee there is a competent teacher in every classroom (Danielson, 2001). As Danielson (2010) points out, “A principal or a superintendent must be able to say to the school board and the public: ‘Everyone who teaches here is good - and here is how I know’” (p. 36). The goal of teacher evaluations that are focused on quality assurance is to aid administrators in understanding if a teacher’s skill is not meeting the current standard and then to use this information to drive employment and compensation decisions (Danielson, 2010). Danielson (2010) points out that this is a critical piece of information in deciding whether or not to grant teachers tenure or whether to promote them to leadership positions.

However, even with the major evaluation changes that occurred after 2009, there have been little changes to the percentage of teachers being identified as unsatisfactory on their evaluations (Kraft & Gilmour, 2017). The increased attention on using evaluations to identify and remove poor performing teachers however will not succeed in improving education if there is not also a focus on creating strong teachers (Darling-Hammond, 2014). This leads to the second primary purpose of teacher evaluations: professional development and support.

It is the nature of teaching itself that makes this second purpose of professional development and support so critical to enhancing the profession and learning of students. The commitment to professional learning is not about a focus on poor teaching that must be fixed. Instead, it is an acknowledgment that because teaching is so difficult there is always room for improvement (Danielson, 2010). Darling-Hammond (2014) argues that what our country needs is a “conception of teacher evaluation systems as part of the teaching and learning system that supports continuous improvement” (p. 5). These systems help ensure that teachers can collaboratively work together to support student achievement instead of being competitively ranked and sorted which undermines learning communities (Darling-Hammond, 2014). Ultimately, when schools design their evaluation system to help provide targeted-supports to teachers that assist them in improving their craft, they work to shift the distribution of effective teaching forward (Connally & Tooley, 2016).

The recent federal initiatives used to prompt the redesign of evaluation systems did in fact acknowledge both primary purposes of accountability and support (Connally & Tooley, 2016). However, the attempts to merge these two ideas has long been a



struggle for states, districts, and schools in their implementation. Danielson (2001), points to the inherent conflict between function of evaluation and the function of coaching which is required to achieve teacher growth and support. To ensure quality, a system must be valid, reliable and defensible while a system designed to promote growth and learning should be more collegial and collaborative (Danielson, 2010). Both Danielson (2010) and Darling-Hammond (2014) emphasize that to ensure both purposes are being met there must be a focus on how the system is created and how the purposes are embedded in the design of the system. Tucker (2014) points to the lack of strategy at the federal level in growing and developing teachers, but according to Connally & Tooley (2016), some success has been found in merging these two ideas. They specifically point to Tennessee as one state where there has seemingly been movement on the use of evaluations for professional growth. As one of the initial winners of the RTTT Grant, Tennessee overhauled its evaluation system and implemented the Tennessee Educator Assessment Model (TEAM) in 2011 (TNDOE Report, 2016). After the fourth year of implementation in 2014-2015, 68 percent of teachers reported that the TEAM model improved teaching. This was up from only 38 percent in 2012-2013 (Connally & Tooley, 2016). In addition, 81 percent of teachers reported that the evaluation system helps them identify areas where they can improve (TNDOE Report, 2016). Although the TEAM system was designed to assure quality and support of teachers, there is competing data questioning whether or not it is meeting both objectives of accountability and support. Specifically, Tennessee's system represents one of several national concerns around teacher evaluation systems.

### **Problems in Implementation of Evaluation Systems**

As stated previously, the two agreed upon primary purposes for teacher evaluations are accountability or quality assurance and professional development or teacher support. These two somewhat conflicting purposes can lead to potential problems during implementation. Darling-Hammond (2014) states, “It is easy for procedures to overwhelm purpose in almost any reform, and this is particularly true for teacher evaluations” (p. 9). In regards to the TEAM system utilized in Tennessee, the Tennessee Department of Education reports that there is a misalignment between evaluation scores and student growth data for 93 percent for teachers whose students’ growth data is placing them in the lowest effectiveness rating of a Level 1. A similar misalignment is shown for 69 percent of the teachers being placed at the second lowest level of effectiveness of a Level 2 (TNDOE report, 2016). This trend in Tennessee is indicative of a problem addressed by Darling-Hammond (2014) regarding states that rely on a test-based metric in conjunction with classroom observations. “Quite often, the two measures do not agree with one another” (Darling-Hammond, 2014, p. 11). Tennessee’s Department of Education has recognized this concerning issue and stated, “The outcomes of evaluation – accountability and improvement – are dependent on having reliable and valid evidence about teacher performance and student learning. Misalignment between observation and individual growth quickly results in mixed messages for educators” (TDOE, 2016, p. 10).

In addition to score misalignment, a second concern regarding evaluations is the previously expressed issue regarding the use of evaluation systems in differentiating the effectiveness levels of teachers. This concern directly impacts the use of evaluations for

accountability. The teacher evaluation reform movement did lead most states to implement more frequent evaluations that were more focused on instruction. However, this has not led to greater differentiation among the performance ratings teachers are assigned (Kraft & Gilmour, 2017). Weisberg et al. (2009) reported that more than 99 percent of tenured teachers were receiving satisfactory ratings with a “fraction of a percent” (p. 11) being ranked as unsatisfactory. In their updated study, Kraft and Gilmour (2017) state this is still the case and “the percentage of teachers rated as unsatisfactory has not changed in the majority of states that have adopted new teacher evaluations systems” (p. 235). However, they do report there is great discrepancy in the percentage of teachers being rated in the category just below proficiency as well as those above it. For example, Georgia reported only 6% of their teachers as above proficient while neighboring Tennessee reported 62% (Kraft & Gilmour, 2017). Ultimately, the majority of teachers continue to be rated in the top rating categories across most states which leaves little ability for evaluators to have the necessary data needed to make personnel decisions (Connally & Tooley, 2016).

Although policy can provide a foundation for expectations, policy alone is not enough for effective implementation (Connally & Tooley, 2016). Policymakers may be shaping parameters, but it is the individual decisions made by evaluators at the school level that are ultimately deciding the percentage of teachers rated in each category (Kraft & Gilmour, 2017). Kraft and Gilmour (2017) identified four areas that are contributing to these individual evaluators rating so few teachers as below proficient. They include time constraints, factoring in teachers’ potential and motivation, personal discomfort

associated with assigning low scores, and the challenges associated with removing teachers.

In addition to the failure of differentiating levels of effectiveness, Tucker (2014) underscores the issues created by the unstated negative assumptions policymakers may hold about evaluation reform. He states that policymakers have continued to place their emphasis on teacher evaluation solely as the way to identify teachers who need to be removed from the profession instead of identifying the needs of teacher development. He contends that policymakers are operating under the assumption that the way to motivate teachers to work harder and more efficiently is to threaten them with loss of their job or public shaming. Connally and Tooley (2017) point out that even though the intent of evaluation reforms was to offer both accountability and support, states have “primarily focused on the former” (p. 4).

### **Elements of Effective Evaluations**

There is consistency in outlining the elements of evaluation systems that have the potential to ensure teacher quality while also providing avenues for development and growth. These include elements that are currently embedded in many state plans such as requiring multiple observations are used, implementing ongoing feedback structures, and ensuring evaluators are consistently well trained (Danielson, 2010; Darling-Hammond, 2014). Connally and Tooley (2017) present these characteristics under the banner of communication, support, and monitoring. They state that effective evaluations begin by strongly communicating the purpose of evaluations as a professional development tool. States and districts must then ensure schools are supported in the practice of connecting evaluation and professional development. Finally, there must be monitoring to ensure

accurate and actionable feedback aimed at professional growth is provided (Connally & Tooley, 2017).

### **Connecting Teacher Evaluations and OKRs**

Doerr's (2018) OKR system used in business provides clear connections with elements of effective evaluations and opportunities to strengthen as the primary purposes of the evaluation system. This begins with the participation of the teacher in the evaluation system. Danielson (2010) points out it is critical that teachers are engaged in the evaluation process to truly promote learning. She pushes back against systems where evaluations are "an activity done to teachers" (Danielson, 2001, p. 14). Instead the emphasis is on engaging teachers in the process and understanding that "learning is done by the learner through a process of active intellectual engagement" (Danielson, 2010, p. 38). Darling-Hammond (2014) frames this engagement through the lens of collaboration and stresses that it is in collegial settings where common goals can be set, and expertise is shared that teaching improves the most. The collaborative system highlighted by Darling-Hammond (2014) is also emphasized in the OKR process. The OKR system underscores the employee's active engagement in the goal setting process through their personal contribution to the objectives and key results (Doerr, 2018). Doerr (2018) emphasizes the need to link together to accomplish their goals and states that when it comes to innovation and advanced problem solving "isolated individuals cannot match a connected group" (p. 89).

The emphasis on feedback and focused conversations is another clear area where OKRs and effective evaluations merge. Darling-Hammond (2014) emphasizes that useful feedback is an element of evaluations that teachers crave, and, as stated earlier,

Connally and Toole (2017) emphasize the importance of ensuring accurate and actionable feedback aimed at professional growth is provided. For OKRs, one way that feedback is provided is through the practice of regular check-ins. On-going check-ins are critical to the OKR process and are “essential to prevent slippage” (Doerr, 2018, p. 117).

Additionally, the OKR system emphasizes that it is through the on-going conversation process where managers evolve from “taskmasters to teachers, coaches and mentors (Doerr, 2018, p. 184). A focus on conversations and feedback not only underscores the importance and value of reflection on professional learning, but it also helps us understand reflection is promoted through dialogue with others (Danielson, 2001).

### **Goal Setting Theory, OKRs, and Motivation**

Doerr (2018) clearly lays out how companies such as Intel, Google, and YouTube grew and profited from the use of the OKR system and how this system provided motivation for employees at all levels of the organizations. The use of goal setting as a motivational tool is found clearly in Locke and Latham’s (1990) goal setting theory which asserts that “task performance is regulated by the conscious goals that individuals are trying for on the task” (p. 240). Doerr (2018) ties Andy Grove’s work to this theory and specifically mentions the connection between specific, difficult goals and stronger performance compared to poorer performance when easy, vague goals are used or when people are simply urged to do their best.

Locke and Latham (2019) emphasize the connection between both the specificity of the goal and the difficulty level and underscore that both attributes are needed to impact performance. Without specificity, goals are subject to interpretation. In addition, the challenge of the goal directly relates to performance and effort given. Therefore, a

specific, easy goal would not lead to higher performance. When individuals set specific goals that are high yet attainable, it not only provides direction for their behavior but also serves as the motivator for performance attainment (Locke & Latham, 2019).

A foundational piece connecting goals and high performance is also found in one's commitment to truly trying to achieve the goals (Locke & Latham, 1990). Locke and Latham (1990) lay out several important practices that are key to increasing motivation and goal commitment including setting attainable goals, publicly committing to the goals, and role modeling of the goal setting process. In addition, feedback is central to the motivational process of goal commitment. Without providing feedback related to tracking one's goals, there is little impact. Further strengthening this connection, providing feedback without a clear goal has little impact on performance (Locke & Latham, 1990). It is the feedback process that provides people the knowledge they need to adjust the level of their actions or increase their effort if needed (Locke & Latham, 2002). Each of these motivational elements is present in the OKR process (Doerr, 2018) which would underscore their importance in creating goals that foster commitment. Locke (2000) emphasizes that goals are a key concept in motivation and therefore directly impact one's action. Not only do goals impact the direction or focus of our actions, but they also directly impact the intensity with which we act and the persistence over time in reaching the goal.

In addition to goals, however, Locke (2000) also points to needs, values, and emotions as three additional key concepts of motivation. Each of these is also significant in the process of the "arousal of action" (p. 411). However, it is the concept of needs that is the starting point for motivation. When considering these concepts, Locke (2000),

indicates that the motivational path travels from needs to values to goals. Understanding that motivation first ties to the fulfillment of one's needs helps connect goal setting theory to Ryan and Deci's (2000) self-determination theory (SDT) and its impact on human motivation.

### **Self-Determination Theory**

At its core, motivation is about being moved to do something. However, Ryan and Deci (2000) emphasize that motivation is not a singular phenomenon, but instead there are various kinds of motivation at play in decision making. The most common distinction is between intrinsic and extrinsic motivation both of which focuses on why actions are taken.

According to Ryan and Deci (2000), intrinsic motivation is in reference to completing things based on its inherent interest or because of its enjoyment. SDT recognizes that humans are curious and inquisitive by nature, and they do not naturally require extraneous incentives to learn and explore (Ryan and Deci, 2000). In addition, people are naturally eager to succeed because success itself is rewarding (Deci & Ryan, 2008a). The benefits of strong intrinsic motivation are clear in that it increases performance, enhances engagement, and is "likely responsible for the preponderance of human learning across the life span" (Ryan and Deci, 2020, p. 2).

In comparison, extrinsic motivation refers to completing an activity because it leads to a separate outcome. Despite the value found in intrinsic motivation, an increasing number of the activities completed post-early childhood are done some for extrinsic reasons (Ryan & Deci, 2000). Deci & Ryan (2008a) point out the dangers of using extrinsic rewards such as money or awards as motivators as they actually decrease



intrinsic motivation and decrease their overall interest in the action itself. They emphasize that one reason for this impact is because extrinsic motivators often feel controlling. However, an important factor in SDT is the understanding that people can feel more autonomous and less controlled while still being extrinsically motivated. Due to this, SDT differentiates various types of extrinsic motivation and categorizes motivation not as simply intrinsic or extrinsic, but as autonomous and controlled motivation (Deci & Ryan, 2008a).

### **Autonomous vs. Controlled Motivation**

Autonomous motivation includes both intrinsic motivation and types of extrinsic motivation in which people can relate to the value of the task and integrate it with their sense of self (Deci & Ryan, 2008b). When people have autonomous motivation, they act of their own volition and self-endorse the task they are completing or working towards (Deci & Ryan, 2008a). In contrast, Deci and Ryan (2008b) point out that controlled motivation refers to behaviors that are done in response to external contingencies such as rewards or consequences. They emphasize that this feeling of being controlled leads them to think and act in specific ways.

The contrast between these two types of motivation are clear and stark. Where autonomous regulation is associated with increased persistence, enhanced performance, greater psychological well-being, better productivity, and less burnout in the work environment; controlled motivation provided through sanctions, rewards, and other external manipulations undermine quality engagement (Deci & Ryan, 2008a). In addition, Deci and Ryan (2008b) found that autonomous regulation can actually be invigorating and build energy where controlled motivation drains energy.

Building on this contrast of autonomous and controlled motivation, SDT further emphasizes that there are basic psychological needs present in people that must be satisfied. These universal needs of competency, autonomy, and relatedness are key to enhancing optimal motivation and supporting people's activities (Deci & Ryan, 2008a). As stated earlier, the idea of needs being a primary and initial motivator for action is also seen in Locke (2000), and it helps us understand how outside forces impact autonomous and controlled motivation (Deci & Ryan, 2008b). An assumption of SDT is that people are inherently motivated towards learning and mastery, and it is in supporting these three basic needs that allow for more robust autonomous motivation (Ryan and Deci, 2020).

#### *Competence (Self-Efficacy)*

The need for competency relates to the need for individuals to feel they can succeed, grow, and obtain a feeling of mastery (Ryan & Deci, 2020). This connection is articulated by Bandura (1997) as the concept of self-efficacy. Like, Deci and Ryan (2008b), Bandura (2000) puts personal self-efficacy as a foundation of human agency, and states, "Unless people believe that they can produce desired effects and forestall undesired ones by their actions, they have little incentive to act" (p. 75). Motivation is impacted and regulated by self-efficacy because people will take stronger incentives when they believe their actions will be effective (Bandura, 1997). Bandura (1997) points out that the beliefs people have about themselves impacts the goals they make and how much effort and perseverance they put forward. In addition, it is self-efficacy that impacts strategic thinking, optimistic thinking, their resilience to adversity, and the accomplishments they realize (Bandura, 2000).

This universal need is not only important to SDT but is a key part of Locke and Latham's (1990) goal setting theory thus further connecting the two theories. People with high self-efficacy consistently set higher goals for themselves than do those with a lower self-efficacy. In addition, high self-efficacy relates to a larger goal commitment and using better strategies to attain the goals (Locke & Latham, 2002). The benefits of having a higher self-efficacy does not end once goals are reached as these people proceed by setting an even higher standard for themselves (Bandura & Locke, 2003). Bandura & Locke (2003) point out the importance of high self-efficacy because the natural process of achieving goals requires a resilient belief that one has what it takes to be successful even in the face of repeated failures and setbacks. In contrast, a lack of self-efficacy or perceived competency leads to an avoidance of difficult tasks (Bandura, 1997). When one has a low sense of efficacy it leads to increased self-doubts, focusing on obstacles, and a tendency to blame their own inadequacies (Bandura, 1997).

In working to build self-efficacy and satisfy the need for competency, feedback again plays a pivotal role in the process. Bandura and Locke (2003), specifically emphasize the value of framing the feedback in terms of gains made towards the goals. When this occurred, people sustained a higher perceived self-efficacy, raised self-set goals, and had higher self-satisfaction. Deci and Ryan (2008a) build on this and points out that although tangible rewards do decrease intrinsic motivation, when people received positive feedback on their performance their intrinsic motivation increased. They point out that positive feedback feeds the sense of competency without conveying a sense of control that is found in other extrinsic motivators. However, not all feedback is equal. Bandura and Locke (2003) and Deci and Ryan (2008a) also both point out the dangers of

negative feedback on personal motivation. When goal feedback is framed as a shortfall, self-efficacy and self-satisfaction both decrease. In addition, future self-set goals decrease. Finally, failing to satisfy peoples' need to feel competent not only undermines their intrinsic motivation, but can leave them amotivated, without any intrinsic or extrinsic motivation (Deci & Ryan, 2008a).

### *Autonomy*

Satisfying the need for competency is critical to increasing motivation. However, SDT emphasizes that competency alone will not enhance intrinsic motivation unless it is accompanied by a sense of autonomy or an "internal perceived locus of causality" (Ryan & Deci, 2000, p. 58). This emphasis on self-determined behavior speaks directly to the juxtaposition of intrinsic versus extrinsic motivation and SDT's autonomous versus controlled behavior. Having your need for autonomy satisfied is ultimately about feeling as if you have choice which in turn will produce higher levels of intrinsic motivation (Deci & Ryan, 2008a). In addition, Pink (2009) outlines that increased autonomous motivation leads to "greater conceptual understanding, better grades, enhanced persistence at school and in sporting activities, higher productivity, less burnout, and greater levels of psychological well-being" (p. 89). Deci and Ryan (2008a) compare this to often used extrinsic motivators including rewards, threats, surveillance, and evaluations that increases the feelings of being controlled and therefore thwarts intrinsic motivation and the benefits that accompany perceived autonomy.

Pink (2009) specifies that autonomous motivation goes against many previous notions of management which treats the employee as "pawns rather than players" (p. 89). This means that leaders and managers play an important role in creating the conditions

that foster the autonomous opportunities and environments. Deci and Ryan (2008a) speak to this as autonomy support, and they point out that it includes creating environments where authority figures take the perspective of those they manage, encourage initiation and choice, and are responsive to their thoughts and questions. Providing this autonomous support has shown to have increasing benefits for both the individual and the organization. One study of a Fortune 500 company found that managers who provided more autonomy-supportive environments had employees who felt less controlled and pressured, were more trusting of management, and reported greater satisfaction with their job and with the work setting (Deci & Ryan, 2008a). Additional studies have shown a correlation to increased worker engagement, evidence of greater worker well-being, and higher worker performance ratings (Baard, Deci, & Ryan, 2004).

It is important to understand that autonomy refers to acting of your own volition and with choice, but it does not mean acting without accountability, acting without structure, or acting independently of others. Pink (2009) again contradicts autonomous support with management styles of the past which he says assumed that providing people with freedom and choice would neglect their responsibility. However, he contends that people must be held accountable and ultimately want accountability, and that “making sure they have control over their task, their time, their technique, and their team is the most effective pathway to that destination” (Pink, 2009, p. 105). In addition, SDT differentiates between the ideas of control and structure. Creating structure includes the setting of clear expectations and goals, ensuring consistency in rules, and providing supports for engagement and feedback which when combined with autonomous support

increases autonomous motivation (Ryan & Deci, 2020). Finally, independence is about functioning alone and not relying on others. When one acts with independence it does not indicate they are always acting autonomously. Instead they can be acting independently while also being controlled. Similarly, one can be having their needs for autonomy met while also being interdependent on others (Ryan & Deci, 2020).

Understanding the difference between autonomy and independence allows us to see why autonomy is not in contradiction to the third universal basic need identified in SDT, a sense of relatedness.

### *Relatedness*

According to Ryan and Powelson (1991), relatedness concerns both the emotional and personal bonds between people. However, they emphasize that it is more than just connection. “Relatedness refers to the experience of connecting with others in ways that conduce toward well-being and self-cohesion in all individuals involved” (Ryan & Powelson, 1991, p. 53). Ryan and Deci (2000) point out that as activities become more extrinsic in nature, the reason people are willing to do things becomes increasingly based on the value they feel by others who are significant to them. This signifies that providing a sense of belonging or connecting people can help internalize extrinsic motivators making them more autonomous (Ryan and Deci, 2000). Relatedness is ultimately connected closely to the other basic need of autonomy. When someone is responsive to the autonomous expressions of others, it actually increases feelings of relatedness (Ryan & Powelson, 1991). Conversely, Ryan and Powelson (1991) report that in a study on students and relatedness, students who experienced controlling adults also reported low security in their relationships. When a person feels involved with a group, an

internalization of values and behaviors are endorsed, and this support will lead to greater intrinsic motivation (Deci & Ryan, 2008a).

Ultimately, Deci and Ryan (2008a) articulate that factors which support high-quality motivation also support the workers' basic psychological needs of competency, autonomy, and relatedness. When this occurs, not only does performance increase but there are also positive impacts on workers' persistence, intensity, and creativity on the job.

### **The Culture for OKRs to Succeed**

One of the primary benefits discussed in the OKR system is that they provide clarity around a leader's priorities. This clarity becomes a vehicle to help to ensure greater performance is achieved for the team. However, clarity alone is not enough, and "goals cannot be attained in a vacuum" (Doerr, 2018, p. 212). The transmission of priorities, insights, and goals requires that the system be implemented in the correct culture. Doerr (2018) builds this bridge by tying OKRs to a performance management system he calls CFR which stands for conversations, feedback, and recognition.

According to Doerr (2018), "CFRs ignite OKRs and then boost them into orbit; they're a complete delivery system for measuring what matters" (p. 176). By emphasizing the importance of conversations, feedback, and recognition in the OKR process, Doerr (2018) is underscoring the importance of creating a culture where this work can thrive. This emphasis on culture is not new in Doerr's incarnation of the OKR system but has its foundation with Andy Grove's work just as the other core OKR components. Grove prized collective accountability, fearless risk-taking, and transparency. Additionally, he also understood the importance trust played in achieving this work which is why he

placed it with the others in his list of core values at Intel (Doerr, 2018). Grove's understanding of the importance of trust is clearly in line with research that emphasizes the critical nature it plays in organizational capacity (Cosner, 2009).

### **A Foundation of Trust**

Research has emphasized many benefits of building trust in the workplace including an increased willingness to disclose information regarding problems and concerns, more cohesive and productive relationships (Tschannen-Moran and Hoy, 2000), increased emotional attachment and commitment to the organization (Cosner, 2009), and a positive impact on attitudinal, behavioral, and performance outcomes (Dirks & Ferrin, 2002). In addition, an individual's trust in a direct leader has an even greater impact on performance, altruism, and job satisfaction than trust in just organizational leadership (Dirks & Ferrin, 2002). Overwhelmingly, trust has become a vital part of well-functioning organizations. At all levels, "trust facilitates productivity, and its absences impedes progress." (Tschannen-Moran & Hoy, 2000, p. 585).

Although there is seemingly wide-spread agreement on the importance of trust, Tschannen-Moran and Hoy (2000), point out the difficulty in defining this complex concept which has been viewed philosophically, economically, individually, and organizationally. In reviewing literature, vulnerability and honesty were found as common themes across multiple contexts, and helped Tschannen-Moran & Hoy (2000) craft the following multidimensional definition of trust: "Trust is one party's willingness to be vulnerable to another party based on the confidence that the latter party is (a) benevolent, (b) reliable, (c) competent, (d) honest, and (e) open" (p. 556).



As previously discussed, key characteristics of the OKR system include setting aggressive goals, making these goals transparent, and frequent check-ins throughout the process (Doerr, 2018). These require a trusting relationship with one's leader because these behaviors rely heavily on risk-taking, vulnerability, and openness, all of which are key elements of trust.

Dirks and Ferrin (2002), emphasize that when others have the sense that their leaders demonstrate characteristics of trustworthiness, they are more comfortable engaging in risk-taking behaviors. This is echoed by Handford and Leithwood (2013) who point out that trust is what enables change to occur, and that it is a critical piece for every interaction in the organization. When new practices are implemented, they will often result in a decrease in performance as the initial efforts are executed. It is trust in leaders that will not only influence one's willingness to risk innovated practices, but it is what will also increase the likelihood that they will temporarily risk unskillful performance (Handford & Leithwood, 2013). Risk taking, therefore, is closely connected to vulnerability because of the implication that there is something to be lost, and it is commonly considered a precondition for trusting relationships (Lapidot, Kark, & Shamir, 2007). Research seemingly points to a reciprocal nature between trust and its key elements. Lapidot et al. (2007) point out that vulnerability will increase an individual's sensitivity to their leader's behavior, and that will then increase the impact the leader's behaviors will have on the growth or decline of trust. Similarly, risk-taking and trust work together in much the same way in that trust allows one the ability to take risk which in turns reinforces a greater sense of trust (Tschannen-Moran & Hoy, 2000).

It is clear that building trust will lead to promising benefits for organizations. According to Tschannen-Moran & Hoy (2000), the responsibility to initiate this trust falls on leaders and demands that they display trustworthy behavior. However, research by Laipdot et al. (2007) emphasizes the complex nature of trust and shows that individuals can experience both trust-building and trust-eroding incidents in their relationships with leaders. This further indicates the importance of leaders being keenly aware of the research-based behaviors associated with trust-building. Whitener, Brodt, Korsgaard, and Werner (1998) point to consistency, integrity, concern, communication, and sharing control as key behaviors needed to cultivate trust. Handford & Leithwood's (2013) research on trust with teachers in school settings also identify consistency and integrity as critical trust building behaviors, but they also include competence, openness, and respect. Not recognizing and embracing these important behaviors will ultimately lead to a decline in trust which will produce outcomes counterproductive to the goals of the OKR process including an unwillingness to take risks, the withholding of information, and increased feelings of insecurity (Tschannen-Moran & Hoy 2000).

### **Trust and Empowerment**

McBride and Skau (1995) state that effective supervision is “built on a foundation of trust” (p. 267), and when in place this foundation can create the conditions needed to empower others. Empowerment is not about having power over others but gaining and exercising one's personal power and increasing one's autonomy (McBride & Skau, 1995). As previously discussed, autonomy plays an important role in the OKR process. First, OKRs emphasize participating in the goal setting process and giving individuals a say in the objectives they are trying to reach (Doerr, 2018). In addition, increasing

autonomy is a key factor in creating work conditions that are more intrinsically motivating (Deci & Ryan, 2008a). Moye, Henkin, and Egley (2004) also connect empowerment, autonomy, and trust in their research which examined the relationship between teachers' sense of empowerment and their interpersonal trust in their principals. Their results indicated that teachers who identified their work as personally meaningful, who reported having autonomy, and who perceived they had influence over their environment also reported high level of interpersonal trust in the principals.

The benefits of creating a culture built on trust that empowers teachers are clear. McBride and Skau (1995) state that empowered individuals are more motivated, understand the need to seek alternative strategies and solutions, and are eager to discuss their practices with their supervisors. In addition, empowered teachers commit to action and accept responsibility for the outcomes. "If the decision does not provide them the desired result, they reflect about the choice made and attempt to learn from the experience" (McBride & Skau, 1995, p.269). In addition, empowerment and trust can have mitigating effects on organizational complexity and diminish the need for oversight and other controls that negatively impact productivity (Moye et al., 2004). Finally, McBride and Skau (1995) point to another important aspect of empowerment that is key to encouraging risk taking, creativity, and responsibility. That is the importance of creating an emotionally safe environment.

### **Psychological Safety**

Psychological safety refers to the "individuals' perceptions about the consequences of interpersonal risk in the work environment" (Edmondson, 2004, p. 241). Through a systematic review of literature, Newman, Donohue, and Eva (2017) point out

there are certain risks involved when employees are encouraged to engage in practices like voicing new ideas, collaborating with others, or experimenting with new ways of doing things. Even though these behaviors may ultimately benefit the organization, they could still lead to situations where individuals need to go against the interest of others, be seen in negative light, or have their risk-taking be considered as a failure. It is clear that these risks could potentially keep employees from contributing to the organization and hinder both the individual's and the organization's learning (Newman et al., 2017).

Garvin, Edmondson, and Gino (2008) argue that organizational learning is critical in today's world of increased competition and advancing technology and that it hinges on employees being comfortable expressing their thoughts on the work being done. "To learn, employees cannot fear being belittled or marginalized when they disagree with peers or authority figures, ask naïve questions, own up to mistakes, or present the minority viewpoint" (Garvin et al., 2008, p. 111). Conclusively, creating environments that are psychologically safe helps to eliminate the threats these risks pose.

Conversely, Edmondson (2004) found that when individuals feel psychologically safe, they are more likely to engage in behaviors such as seeking feedback, seeking help when needed, speaking up about concerns and mistakes, innovating, and expanding boundaries. In addition, research points to an increase in individual reflection and working around issues when encountering blocks in the workflow (Newman et al., 2017). All of these are not only behaviors that will increase the likelihood of organizational learning, but they are all also behaviors important to the OKR process (Doerr, 2018).

Similarly, to Tschannen-Moran & Hoy's (2000) viewpoint on trust, Edmonson (2004) emphasizes the important role leaders plays in enhancing the perceptions of

psychological safety in the workplace beginning with leaders ensuring they are accessible, open, and available to their team. In addition, leaders must not only encourage individuals to voice ideas and take strategic risks, but they must communicate that these behaviors are appreciated and respected (Page, Boysen, & Arya, 2019). Page et al. (2019) illustrates that when this happens, we again see a reciprocal nature between psychological trust and desired employee behaviors. “When a leader displays openness to hearing employees’ ideas and concerns while actively listening to new ideas that are brought forth, employees will feel comfortable, and in return be more open and willing to share creative ideas to meet challenges” (p. 30.). However, perceived feelings of judgment will result in the holding back of opinions and ideas (Edmondson, 2004).

Edmondson (2019) further clarifies the roles leaders can play in increasing psychological safety by reframing failure in the organization. In her work, preventable failures, which are never good, are considered separate from complex and intelligent failures which can provide important information. Furthermore, intelligent failures can even be considered good news because of the increased value they bring. Edmondson (2019) emphasizes, however, that understanding the important information failure can provide must be accompanied with enough psychological safety in the organization to allow individuals to dig into those lessons.

Although psychological safety is important in all organizations, Edmondson, Higgins, Singer, and Weiner (2016) specifically point to the importance of it in high-stakes and challenging environments such as education. As they point out, teachers are generally responsible for 20 – 30 students who have varied skills and backgrounds. This leads to a demand for customization, improvisation, and individual attention creating a

context where psychological safety is most valuable (Edmondson et al., 2016). Their research again pointed to the importance of the leader and found that “workers’ perceptions of their leaders’ ability to facilitate positive work environments affects psychological safety” (p. 78).

Ultimately, trust is a critical foundational piece of an organization that seeks to empower their employees and provide the psychological safety needed to take risk, innovate, and work interdependently with others to reach individual or group goals at the center of the OKR process. As Doerr (2018) states, “Healthy culture and structured goal setting are interdependent. They’re natural partners in the quest for operating excellence” (p. 213).

### **Expertise**

A distinguishing factor of the OKR system is the consistent connection between an organization’s objectives and the key results that will lead to accomplishing them. As previously discussed, Grove’s emphasis on key results is what set Intel apart from other companies who tended to value experience regardless of whether or not that led to results. This emphasis on execution laid the foundation for the OKR system (Doerr, 2018). The delineation between experience and execution can also be viewed through the lens of differentiating experience from the building of expertise in your practice, an area that has been of increasing interest to researchers since the late 1800s (Ericcson, Krampe, & Tesch-Romer, 1993).

Ericcson (2018a) provides several defining characteristics for what it means to be an expert including that they are skillful, well-informed, and recognized, often publicly, for their knowledge, technique, and skill. Therefore, having expertise in an area refers to

displaying the “characteristics, skills, and knowledge that distinguishes experts from novices and less experienced people” (Ericsson, 2018a, p. 4). Dall’Alba (2020) emphasizes that being an expert, however, is more than just having increased knowledge or skills. Inherent to the ideas of expertise is the expectation of high performance in the field and the ability to consistently display a higher-level of performance than most others (Dall’Alba, 2020). The building of expertise in an organization can therefore be an important part of the improvement process. However, organizations must understand how expertise is developed to ensure the expected high-performance is realized.

Early understandings around expertise often centered around the belief that individuals were genetically predisposed to innate talents which lead to higher performance (Ericsson et al., 2013). This idea was not new, and as Ericsson and Charness (1994) point out, since the emergence of civilization many believed that desirable, individual attributes were “gifts from the gods” (p. 726). However, current research around the development of expertise has largely rejected the importance of innate ability (Ericsson & Charness, 2014). Ericsson et al. (2013) emphasizes that their review uncovered “essentially no support for fixed innate characteristics that would correspond to general or specific natural ability and, in fact, has uncovered findings inconsistent with such models (p. 399).

Ericsson et al. (2013) do specifically point out that better organized knowledge and skills needed to reach expert performance must be acquired. However, they discredit the belief that prolonged engagement in activities or enough experience or practice leads to this maximum performance. In contrast, studies have shown that practice and experience move individuals only to a certain point of acceptability. After that, their

abilities appear to plateau without further improvements being observed (Ericsson, 2018b). Further emphasizing this point, the performance of highly experienced individuals is often comparable to that of less experienced individuals in fields such as nursing (Ericsson, 2018b). Choudhry, Fletcher and Soumerai (2005) found that in the nursing field, experience has even shown reductions in performance which they associate most likely with forgetting. Here we see that continuous improvement is not an automatic byproduct or consequence of more experience. Instead, specific kinds of practice or training task must be pursued for those seeking expert achievement (Ericsson, 2018a).

### **Deliberate Practice**

Although Ericsson et al. (1993) showed that maximum performance is not automatically attained through extended experience, they did show that performance can be increased, even by highly experienced individuals. This increase is found when individuals participate in deliberate efforts to improve. Deliberate practice is characterized by distinguishing between experiences that are designed for the primary purpose of improving skills from the mere participation in activities where learning may be an indirect result (Ericsson et al., 1993). Deliberate practice has been defined as “the individualized training activities specifically designed by a coach or teacher to improve aspects of an individual’s performance through repetition and successive refinement” (Ericsson & Lehmann, 1996, p. 278-279). Unlike practice or engagement alone, deliberate practice has been found to be most effective in improving performance (Ericsson et al., 1993), and has shown to have a significant positive relationship with both objective and relative performance (Ericsson & Harwell, 2019).



Ericsson et al.'s (1993) original research around deliberate practice focused mainly on the domains of music and chess. However, the research laid out specific characteristics that will allow us to transfer deliberate practice to other domains of expertise. These characteristics include one-on-one instruction by a coach or teacher, the assigning of practice activities with explicit goals and immediate feedback, and the opportunity for repetition (Ericsson, 2018b). In addition, these characteristics also begin to show an overlap with Doerr's (2018) OKR system which also focuses on objectives that are reached through intentional key results and frequent opportunities for feedback.

### **Characteristics of Deliberate Practice**

The first characteristic that separates deliberate practice from other types of less effective practice is the presence of a well-qualified teacher in the process who contributes to the learning in ways that cannot be obtained in other types of practices where a teacher is not present. This includes giving explicit instruction about new or best methods, the individual diagnosis of errors, informative feedback, and remedial part training (Ericsson et al., 1993). In addition, the teacher can communicate the goals to be achieved and is able to assess specifics about where an individual can improve between meetings. This allows for the design of specific practice techniques to maximize learning (Ericsson & Harwell, 2019). These intentionally designed, teacher-led practice activities are also what separates deliberate practice from other types of engagement-based activities such as play and traditional work. Play, in contrast to highly structured deliberate practice, includes activities that have no specific goals (Ericsson et al., 1993). Research by Ward, Hodges, Starkes, and Williams (2007) confirm the disconnect between play and expertise, and they found that those achieving elite status in the

domain of soccer spent less time in playful activities between the ages of 14 and 18 than sub-elite players.

Traditional work includes activities that are motivated by external social and monetary rewards instead of motivation that is tied to the goal of improving performance, and although learning can occur in a work environment, it is not the purpose of the activity (Ericsson & Charness, 1994). The traditional work definition given previously also demonstrates another connection between deliberate practice and the OKR process. In contrasting work with deliberate practice, Ericsson et al., (1993) point out that traditional work is characterized by individuals relying on proven methods in an attempt to minimize the cost associated with mistakes and failures. This cost “discourages learning and acquisition of new and possibly better methods during the time of work” (Ericsson et al., 1993). However, both OKRs (Doerr, 2018) and deliberate practice (Ericsson et al. 1993) focus attention on intentional goals specifically designed to improve performance. In addition, both allow for experiences where individuals can attend to specific aspects of their performance driven in part by feedback from their teacher or leader.

The provision of feedback to individuals is central to deliberate practice just as it is to the OKR process. Ericsson et al. (1993) emphasize that individuals need immediate feedback on their progress and knowledge of the results on their performance. They state that “in the absence of adequate feedback, efficient learning is impossible and improvement only minimal even for highly motivated subjects” (Ericsson et al., 1993, p 367). As mentioned previously, the teacher is critical in communicating the goal to be achieved for improved performance. Ericsson and Harwell (2019) point out that it is the

feedback given by the teacher regarding how well-stated goals were attained that will help individuals both refine their mental representations and notice differences that have yet to be achieved. In this way, feedback is the key to moving towards improvement and expertise. Without feedback, the performance of the same activities over and over not only results in the plateauing of performance as discussed earlier, but it also leads to a reduction in the level of effort given (Ericsson, 2018b). Ultimately, Ericsson (2018b) points out that not all types of practices are equally effective in improving performance. Improvement is found when individuals engage in the types of deliberate practice discussed above which are led by a skilled teacher and integrated with on-going feedback.

### **Deliberate Practice and Motivation**

The benefits of deliberate practice on improved performance are clear. However, Ericsson et al. (1993) emphasize that engagement in deliberate practice is not inherently motivating. In lieu of being motivated by external rewards, the motivation for participating in deliberate practice is found in its connection to improved performance therefore making it an effortful activity that requires intentionality and focus. A study of expert and intermediate football players by Coughlan, Williams, McRobert, & Ford (2013) confirms Ericsson et al. (1993) findings. Coughlan et al.'s (2013) research found that expert football players rated practices as more effortful and less enjoyable than their intermediate counterparts. In addition, the expert players chose to practice their weaker skills significantly more than the intermediate group who focused more on practicing their stronger skills. Further confirming Ericsson et al. (1993) theory, the expert players used available feedback more effectively and showed a relatively permanent

improvement in their practice which did not occur for the intermediate group (Coughlan et al., 2013).

In addition to the less enjoyable nature of deliberate practice, the practice must also be sustained over an extended period of time (Ericsson et al., 1993). Gladwell (2008) popularized the notion of extended practice in his writings on the 10,000-hour rule in which he proposed that 10,000 hours was the minimum amount of time necessary to reach expertise. Although other research has confirmed the need for extended period of training and practice, they have not confirmed evidence of a specific number (Ericsson & Harwell, 2019). In fact, Ericsson and Harwell (2019) indicate it could actually be much higher than 10,000 hours in some cases. Ericsson et al. (1993) has shown however that highest level of performance and achievement are often found after approximately 10 years of prior preparation further indicating that the maximization of deliberate practice is not a short-lived process. This connection between time and performance is further explained in the “monotonic benefits assumption” which states, “the amount of time an individual is engaged in deliberate practice activities is monotonically related to the individual’s acquired performance” (Ericsson et al., 1993, p. 368).

Because participation in deliberate practice is neither inherently motivating nor enjoyable, one must consider why individuals elect to participate in it. One answer may be found in Ward et al.’s (2007) study on deliberate practice with elite and sub-elite soccer players. They found that the elite players perceived themselves to be more competent than the sub-elite players and therefore were more likely to invest the time in the challenging task. They further emphasize that the investment in time leads to increased ability levels which continues to increase the players’ perceived competence.

This cyclical idea between deliberate practice, competence building, and intrinsic motivation supports Bandura's (1997) previously discussed research on self-efficacy which states people will take stronger incentives when they believe their actions will be effective. Ericsson et al. (1993) clearly points out that the constraints of effort, time, and motivation play a significant role in the deliberate practice process. Ultimately, they conclude that it is the commitment to deliberate practice that eventually distinguishes the expert from the vast majority of others who cannot meet these demands.

### **Deliberate Practice and Teaching**

Although the majority of research around deliberate practice has focused on practices like sports, chess, and music, the goals of improvement are found in all domains and are especially crucial in the area of teaching. Willingham (2009) reminds us that "teaching, like any complex cognitive skill, must be practiced to be improved" (p. 189). Research by Dunn and Shriner (1999) specifically looked at the impact of deliberate practice in teaching to identify areas that could enhance performance. In their study, they identified six practices that teachers engage in that reflect the characteristics of deliberate practice. These include preparing materials, mental planning, evaluating written work, informally evaluating students, written planning, and formal evaluation of students through self-made tests. Much like the difference between practice and deliberate practice, Dunn and Shriner (1999) found that participation alone did not guarantee learning but that certain activities had the potential for providing opportunities for teachers to learn. Consistent with other deliberate practice activities outside of teaching, engaging in planning and evaluation were not found to be inherently enjoyable. However, for the motivated teacher, they provide "unlimited opportunities to acquire new

knowledge of teaching and to incorporate this knowledge into their unique repertoire of teaching strategies” (Dunn & Shriner, 1999, p. 644).

Dunn and Shriner (1999) emphasize however the innate differences between teaching and the other domains where deliberate practice has been studied. They point out that to judge the effectiveness or improvement of teachers we must look not at the teacher’s behaviors but to a change in the student’s behavior. In other words, “teachers do not ‘practice’ teaching in order to improve but instead engage in patterns of planning, evaluation, and revision so that students improve (Dunn & Shriner, 1999, p. 647). This is supported by DuFour et al. (2016) who also emphasize that the goal in education is to improve student learning. In the end, Dunn and Shriner (1999) found that the deliberate practice framework provides a useful approach to the development of teaching expertise. However, they emphasize that perhaps it is better to consider that “deliberate practice for teachers is approaching normal activities of teaching in a ‘deliberate’ way” (p. 647).

Ultimately, deliberate practice and the OKR system share the same key characteristics of intentional, goal-driven performance that is fostered through coaching and on-going feedback. This connection, and the relationship both have to increasing intrinsic motivation, provides an avenue to bridge the OKR system a relevant improvement method for educators.

The preceding review of literature clearly shows that implementing an OKR system in schools can provide a purposeful avenue to maximize school improvement efforts. By building off of Locke and Latham’s (1990) goal setting theory and Ryan and Deci’s (2000) Self-Determination theory, ORKs not only help to increase teacher efficacy, but they also ensure teachers are focused on results. This will in return increase

the likelihood of educators embracing deliberate practice to show improvements in the teaching and learning occurring in their buildings.

## **CHAPTER III: METHODOLOGY**

### **Introduction**

In an effort to improve schools, both federal legislation and educational research (Hattie, 2012, Darling-Hammond & Goldberg, 2001) have pointed to the vital role that teachers play in enhancing educational progress. Therefore, it is critical that schools consider improvement initiatives that include a focus on improving teacher performance while also honoring the complex nature of the teaching profession. Schools should seek to provide improvement avenues that allow educators to set clear objectives, obtain feedback, and clearly measure their success.

It is clear that teacher improvement must be a substantial part of the school improvement process. However, with so many reform initiatives at play, there is the danger that teachers may be unclear on what improvements they should be striving for and what success in the area looks like. If improving teacher practice is the key to improving schools and student learning, it is imperative they are provided a clear roadmap on how to improve. The preceding problem associated with school and teacher improvement is the foundation for this research study.

One improvement method found in the business community is based off the work of John Doerr (2018) as detailed in his book *Measure What Matters*. Doerr (2018) provides a context around the implementation of an OKR (Objectives and Key Results) system used in various businesses and details the success these businesses have seen because of the implementation.



The purpose of this quasi-experimental study is to seek to collect quantitative data to evaluate and inform on the impact of applying the OKR process as a method for school improvement. In this study, I will investigate improvements as measured by both teacher improvement, student academic improvement, and teacher motivation improvement to answer these questions:

1. What impact does the creating and monitoring of individual, OKR goals have on teacher performance as measured by teacher evaluation data?
2. What impact does the creation and monitoring of individual, OKR goals have on student learning in the English/language arts content area as measured by student performance data?
3. What impact does the creation and monitoring of individual, OKR goals have on student learning in the mathematics content area as measured by student performance data?
4. What impact does the creation and monitoring of individual, OKR goals have on teacher motivation as measured by perception survey data?

### **Research Setting**

The setting for this research will be contained to Southwest Elementary, a kindergarten through sixth grade, suburban public elementary school located in the state of Tennessee. Southwest Elementary is a new school located within Central Schools District, and it will be launching its second year as a school community when the research study begins. Southwest Elementary is one of thirteen schools located in the Central School District, and both Southwest Elementary and the district have a diverse student population. Southwest ended its first year with approximately 807 students, and

table 2 outlines the combined percentage of Black, Hispanic, and Native American (BHN) students that Southwest serves as well as the percentages found in Central School District and in the state. For the purpose of accountability, Tennessee combines these three subgroups as they have been historically underserved populations within the state.

Table 2

*Percentage of Black, Hispanic, and Native American (BHN) Students*

<b>Southwest Elementary</b>	<b>Central School District</b>	<b>State</b>
33.9%	45.2%	35.2%

Table 3 displays the percentage of students who are certified as economically disadvantaged at the school, district, and state level. It is important to note that the percentage of students listed as economically disadvantaged only includes students who receive some sort of direct services related to income and meet the state criteria. There are a number of students who currently qualify for free or reduced lunch in the school setting but who do not meet state criteria to be considered economically disadvantaged. The percentage of students meeting the state definition for economically disadvantaged therefore is not synonymous with students who receive free or reduced lunch.

Table 3

*Percentage of Economically Disadvantaged Students*

<b>Southwest Elementary</b>	<b>Central School District</b>	<b>State</b>
21.3%	32%	34.9%

As demonstrated in the tables above, Southwest currently has a lower percentage of minority students and economically disadvantage students when compared to both the district and state demographics. While the percentage of BHN students is significantly lower than the district, the percentage does closely mirror state percentages. However, the percentage of economically disadvantaged students found in the school is significantly lower than both the district and state.

Southwest does not have current state-wide achievement data to compare their performance to that of other schools in the district or state. In addition to finishing their first year as a school, the COVID-19 pandemic forced school closures during the spring of 2019 which resulted in no state testing.

### **Methodology**

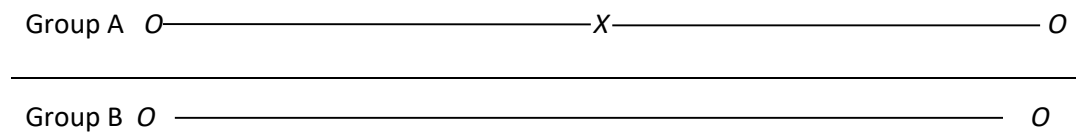
To understand the impact the implementation of OKRs can have on student achievement and teacher performance, a quantitative approach that reflects a Postpositivist worldview will be used by the researcher. Creswell and Creswell (2018) emphasize that this is a connection between quantitative studies and the Postpositive view when researchers seek to examine the relationships between variables in a way that they can be measured and support the idea that there are causes that work to determine effects and outcomes in situations.

A primary goal for the research design is to make inferences about the impact of OKRs in the school setting and determine if there are results that can be generalized to a larger population. Based on this goal, an experimental design was considered. Creswell and Creswell (2018) state that experimental designs “systematically manipulate one or more variables in order to evaluate how this manipulation impacts an outcome (or

outcomes) of interest” (p. 147). They further state that the use of participant random assignment within the study is a principal distinguishing feature that separates experimental designs from other quantitative studies such as a survey study (Creswell & Creswell, 2018). Goplan, Rosinger, and Ahn (2020) refer to the use of randomized controlled studies as the “gold standard” (p. 219) because when it is used “we can confidently claim that the treatment is the most plausible driver of the outcome (p. 219). However, in this research study the participants will be limited to teachers found within the school setting being studied. Considering that the researcher will have only partial control over the random assignment of participants, a quasi-experimental design will be used instead of an experimental design. Like experimental designs, quasi-experimental designs use control and experimental groups and help improve our understanding of the causal effects of practices (Goplan et al., 2020). Although Goplan et al., (2020) underscore the value of true experimental designs, they do also emphasize that quasi-experimental designs are often the appropriate choice in education settings.

The figure below illustrates the research design to be used in the quasi-experimental study. In this diagram, O represents the recording of beginning of year and mid-year student achievement data and teacher evaluation comparison data. The X represents the exposure to the experimental, independent variable (Creswell & Creswell, 2018).

Figure 1

*Nonequivalent Control-Group Design*

The table below outlines the major steps in the research process including the selection of participants, the collection of data, and the implementation of the OKR process for participants.

Table 4

*Overview of Steps in the Research Procedure*

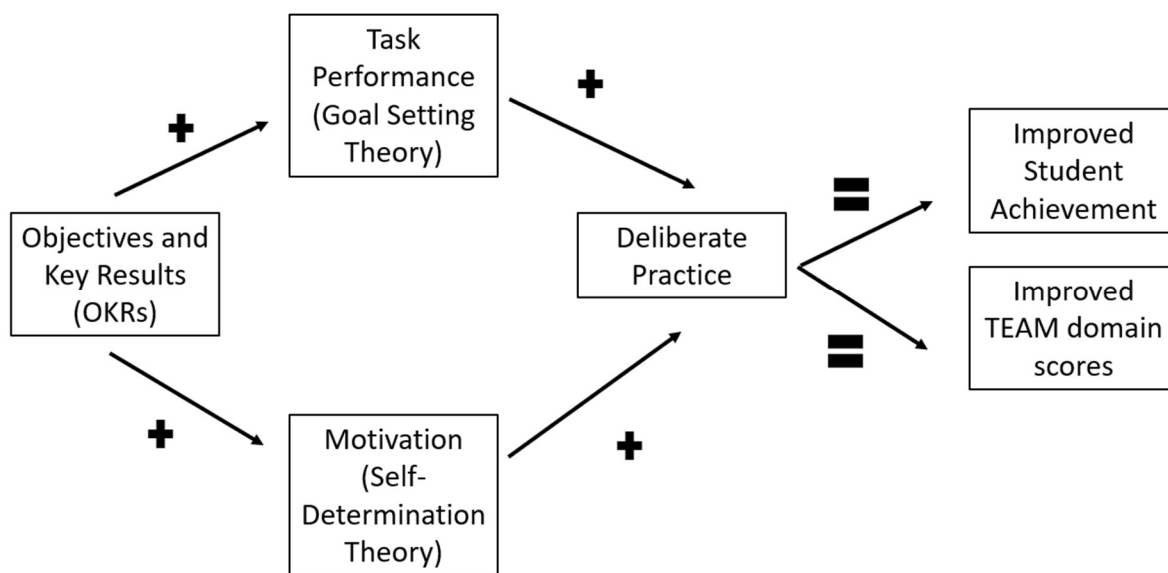
<b>Step Number</b>	<b>Step Description</b>
1.	Teachers in both the control and experimental group will administer the beginning of the year CASE English Language Arts assessment and math i-Ready Diagnostic assessment to all students.
2.	Teachers will be randomly selected and invited to participate in the study thus forming the experimental group.
3.	Teachers in the experimental group will receive training on the OKR process and on how to write OKRs.
4.	Teachers in the experimental group will write at least one individual OKR for the first semester.
5.	Teachers in the experimental group will meet with researcher to review their OKR and identify a connection to a specific performance indicator between their OKR and the TEAM instructional domain rubric.  Teachers in the control group will self-select a specific performance indicator from the TEAM instructional domain rubric that that they would like to improve on based on previous feedback or past performance.
6.	Teachers in both the control and experimental groups will participate in the district's TEAM evaluation model including formal feedback sessions as outlined by state and district policy.

7. Teachers in both the control and experimental groups will administer the mid-year CASE English Language Arts and math i-Ready Diagnostic assessments to all students.
  8. Teachers in both the control and experimental groups will be invited to participate in an anonymous perception survey to answer questions regarding teacher motivation.
  9. Results will be analyzed using an independent samples T-Test to determine if there is a statistically significant difference between the control and experimental groups in regard to student achievement and teacher performance.
- 

## **Rationale**

Throughout his work, Doerr (2018) lays out the impact the implementation of OKRs has had on a wide range of private sector businesses such as YouTube, Google, Intel, and Intuit as well as philanthropic organizations such as The Gates Foundation. OKRs are specifically credited as being the “scaffolding for Google’s signature home runs, including seven products with a billion or more users apiece: Search, Chrome, Android, Maps, YouTube, Google Play, and Gmail” (Doerr, 2018, p. 14). As laid out in the literature review of chapter two, the structure of OKRs is embedded with key practices that support autonomy, goal setting theory, self-determination theory, and deliberate practice. Research on each of these has shown correlations to improved performance. Figure 2 below provides an overview of the logic model connecting OKRs to improved student achievement and improved TEAM evaluation domain scores.

Figure 2

*Logic Model*

As schools across the county continue to look at improvement efforts, this study will provide context around a specific method that has the potential to make significant improvements within the school or district. Doerr (2018) refers to OKRS as a “launch pad” (p. 245), and he specifically states that it has the potential to have a large impact on school success.

Implementing OKRs with the experimental group will assist in determining if the treatment influences the educational outcomes discussed above. Due to this, the quasi-experimental design will be used. Goplan et al. (2020) emphasize that quasi-experimental designs “have significantly improved our understanding of causal relationships in education. Specifically, these methods have been integral in highlighting the strength and



magnitude of causal effects of key educational policies and interventions on well-defined student outcomes” (p.237).

The research-based theories supporting the practice in conjunction with the low financial cost associated with implementation make it an important improvement method that has yet to be studied as a potential school improvement method.

### **Participants and Data Sources**

The participants for this study will be pulled from the population of 45 kindergarten through 6<sup>th</sup> grade classroom teachers at Southwest Elementary School. The teaching staff of Southwest Elementary also includes other certified teachers such as special education teachers, music and art teachers, and intervention specialists. However, only teachers who are currently full-time classroom teachers will make up the population. This will allow for a more precise connection between OKRs and the student achievement data. From this population a random sampling of teachers will be generated. These randomly selected teachers will be invited to participate in the study. The intent will be to create an experimental group that includes 22 educators from the building which will represent half of all classrooms. This will provide an equal number of teachers in our control group and experimental group.

The experience level of the classroom teachers at Southwest Elementary ranges from zero years of experience to 19 years of experience. The average experience is 7.20 years with 20 of the 45 teachers having less than five years of experience. Due to the addition of growth position at Southwest and changes in other staff, 16 of the 45 teachers will be starting their first year as a teacher at the school.

To protect the identity of all participants, the school and district name is reported as a pseudonym. In addition, all participants will be assigned numbers during the reporting process to protect their individual identify. All participants are kept anonymous and no individual student names or scores will be used in the study.

The following four data sources are described in more detail below: CASE English Language Arts assessment data, i-Ready Diagnostic math assessment data, TEAM teacher evaluation data, teacher perception motivational survey. All students in kindergarten through 6<sup>th</sup> grade will participate in the i-Ready Diagnostic math assessment, and all teachers will participate in the TEAM teacher evaluation system. However, only students in 2<sup>nd</sup> through 6<sup>th</sup> grade will participate in CASE English Language Arts assessment.

#### *Case English Language Arts Assessment*

The Case English Language Arts assessments are administered in Central School District three times per year to gauge the academic progress of students within the English language arts content area. The first administration is done at the beginning of the academic year. A second administration is completed at the conclusion of the first semester and a final administration is completed at the end of the academic year. Assessment questions are designed to mirror those of end-of-year state test and reports are generated to help teachers assess where students are in their learning and to make instructional decision to help target instruction (TE21, Inc. CASE Benchmark Assessments n.d.). The assessment is given in a paper/pencil format and students with special testing accommodations as outlined in individualized education plans (IEPs) or

federally recognized 504 plans are provided the appropriate accommodations during testing.

The CASE assessment system seeks to increase reliability by providing two or more equivalent forms of the test that are being administered throughout the year. Tests are untimed and contain multiple questions over individual standards. CASE reports that the assessments typically show above a 90% predictability on how students will perform on state assessments (TE21, Inc. CASE Assessments Research Overview n.d.).

Assessment results are provided to teachers that show the difficulty level of each question based on the P-value of correct answers as well as how each student performs on each individual question. Reports provide a percent correct score for each student as well as projected achievement level.

#### *i-Ready Diagnostic Math Benchmark Assessment*

The i-Ready Diagnostic math assessment is administered in Central School District three times per year to gauge the academic progress of students within the mathematics content area. The first administration is done at the beginning of the academic year. A second administration is completed at the conclusion of the first semester and a final administration is completed at the end of the academic year. This assessment is an adaptive assessment that is administered online through the i-Ready website. The assessment is designed to show students' performance at the time of the assessment as well as growth from the previous diagnostic assessment (Adaptive Assessment That Dives Deep, n.d.). The Central School District uses this assessment as a measuring tool, but it also provides students with an individualized online learning path

based on their results. Due to this, individualized accommodations are not provided. However, the test is untimed.

i-Ready is produced by Curriculum Associates, and they partnered with the Educational Research Institute of America (ERIA) to conduct large scale studies to provide evidence of test validity. The state specific research measured the correlation between the i-Ready Diagnostic and the 2016-2017 end of year assessment in Tennessee, TNReady. The report found a mathematics average correlation of .83 when looking at 3<sup>rd</sup> – 8th grades (Curriculum Associates, n.d.). Field (2018) references the widely used Cohen's *d* benchmarks of 0.2 (small), 0.5 (medium), and 0.8 (large) when discussing effect sizes. According to these benchmarks, the .83 correlation between the i-Ready Diagnostic and TNReady demonstrates a large effect size and supports the high reliability of these scores. This high statistical reliability coupled with the district's affirmation of its value by requiring it of all students strongly suggest this is a reasonable metric to include.

Assessment results are provided to teachers that show performance at both the student and class level. Individual scale scores for each student are provided as well as data indicating increases or decreases in scale scores from previous assessments given that year. Teachers are also provided an approximate grade equivalency for each student in each math domain assessed. The reports also provided typical and stretch growth goal benchmarks for each student based on their scale score.

#### *Tennessee Educator Assessment Model (TEAM)*

The TEAM evaluation model was developed in collaboration between the National Institute for Excellence in Education (NIET) and the Tennessee Department of

Education (TDOE) based on a variety of research on teacher accountability systems, teacher performance competencies, various state specific instructional guidelines and standards, and teacher accountability rubrics (TDOE, 2011). The TEAM model is a state-approved model school districts use to fulfill requirements around teacher evaluation. Central School District has used the TEAM evaluation system as their primary teacher evaluation tool since it was released during the 2011-2012 school year.

This rubric-based evaluation system is broken down into four specific domains with each domain containing specific performance indicators. Table 5 below displays each of the four domains and the performance indicators evaluated in each domain (TDOE, 2018).

Table 5

*TEAM Evaluation Domains and Performance Indicators*

Planning	Instruction	Environment	Professionalism
1. Instructional plans	1. Standards and objectives	1. Expectations	1. Professional Growth and Learning
2. Student Work	2. Motivating Students	2. Managing Student Behavior	2. Use of Data
3. Assessment	3. Presenting Instructional Content	3. Environment	3. School and Community Involvement
	4. Lesson Structure and Pacing	4. Respectful Culture	4. Leadership
	5. Activities and Materials		
	6. Questioning		
	7. Academic Feedback		
	8. Grouping Students		
	9. Teacher Content Knowledge		
	10. Teacher Knowledge of Students		
	11. Thinking		
	12. Problem Solving		

When being evaluated on a specific rubric, teachers receive scores ranging from 1 to 5 for each performance indicator. Although teachers may be assigned a score of 2 or 4 on any performance indicator, performance definitions are only provided for a score of 5 (significantly above expectations), 3 (at expectations) and 1 (significantly below expectations). All evaluators must be certified each year before beginning the evaluation process by completing an online certification test (TDOE, 2018).

Teachers participate in various numbers of observations based on their licensure status and previous evaluation scores. State guidelines allow for multiple domains to be observed during a classroom visit, however, at least half of all observations must be unannounced. The figure below obtained from the TEAM-TN website (<https://team-tn.org/wp-content/uploads/2013/08/Observation-Guidelines-2018-19.pdf>) outlines the specific number of evaluations for each teacher. As outlined, all teachers will be observed on multiple domains each semester.

Figure 3

*TEAM Required Evaluations*

Educator Licensure Status <sup>1</sup>	Previous Individual Growth or Level of Overall Effectiveness (LOE) <sup>2</sup>	Minimum Required Observations	Minimum Required Observations per Domain	Minimum Number of Minutes per School Year
Practitioner <sup>3</sup>	Levels 1-4	Six (6) domains observed with a minimum of three (3) domains observed in each semester.	3 instruction 2 planning 2 environment	90 minutes
	Level 5	One (1) formal observation covering all domains first semester; two (2) walk-throughs second semester	1 instruction 1 planning 1 environment	60 minutes
Professional	Level 1	Six (6) domains observed with a minimum of three (3) domains observed in each semester.	3 instruction 2 planning 2 environment	90 minutes
	Level 2-4	Four (4) domains observed with a minimum of two (2) domains observed in each semester	2 instruction 1 planning 1 environment	60 minutes
	Level 5	One (1) formal observation covering all domains first semester; two (2) walk-throughs second semester.	1 instruction 1 planning 1 environment	60 minutes

Note: An LEA or charter school using the TEAM model may choose to allow observers to combine domains during classroom observations provided the requisite minimum time, semester, distribution, and notice (announced vs. unannounced) are met.

*Teacher Motivation Survey*

The teacher motivation survey used in this study was created by selecting eight questions from the 2020 Tennessee Educator Survey (TES). This survey has been administered yearly since 2011 to public school teachers in the state of Tennessee (Tennessee Educator Survey, n.d.). The complete Tennessee Educator Survey was

developed by the Tennessee Department of Education in conjunction with the Tennessee Research Alliance, and the purpose is to provide educators the opportunity to provide information on what is working in Tennessee's education system and what improvements can be made. The data generated from this survey provides insight to influence state, district, and school level goals (Tennessee Education Research Alliance, n.d.).

The eight questions selected to be used on this study's survey were identified based on their alignment to teacher motivational factors relevant to the study. As in the TES, the questions allow for teachers to rank their responses in one of four possible categories: strongly agree, agree, disagree, strongly disagree. The Tennessee Department of Education has affirmed the value of the TES by consistently using this tool in their data collection process. Considering the eight questions on the study specific survey were drawn directly from the TES strongly suggests that this is a reasonable metric to use.

Table 6 below provides an overview of the independent and dependent variables used for each research question in the study.



Table 6

*Independent and Dependent Variables by Research Question*

Research Question	Independent Variable	Dependent Variable
RQ1: What impact does the creating and monitoring of individual, OKR goals have on teacher performance as measured by teacher evaluation data	Creation and monitoring of individual, OKR goals	TEAM instructional domain average score growth and identified performance indicator growth
RQ2: What impact does the creation and monitoring of individual, OKR goals have on student learning in the English/language arts content area as measured by student performance data?	Creation and monitoring of individual, OKR goals	Percent correct growth for CASE English Language Arts assessment
RQ3: What impact does the creation and monitoring of individual, OKR goals have on student learning in the mathematics content area as measured by student performance data?	Creation and monitoring of individual, OKR goals	Scale score growth on the i-Ready Diagnostic assessment
RQ4: What impact does the creation and monitoring of individual, OKR goals have on teacher motivation as measured by perception survey data?	Creation and monitoring of individual, OKR goals	Responses indicated on the motivation survey.

**Data Collection Procedures**

Data collection for the study began immediately following the researcher obtaining IRB approval for the study and approval from the Central School District.

The CASE English Language Arts assessment and the i-Ready Diagnostic are administered to students by school staff three times a year in the fall, winter, and spring. The exact testing windows are developed by district staff in the central school office. All schools in the Central School District take the assessments during the same window. Assessment data from both student assessments will be collected following the fall and winter administration.

After the completion of the CASE English Language Arts assessment, test and answer documents are collected and shipped to the testing vendor. The testing vendor scores the assessment and generates all reports. Reports are made available within 48 hours of assessments being received and scanned. Copies of school and district reports are made available to the school administration through the vendors secure online website.

Students are administered the i-Ready Diagnostic through an online platform. Individual student reports are available to teachers and school administrators immediately following the completion of the test. Classroom reports are available when all students have tested or at the end of the district set testing window. Since the researcher is also the administrator at Southwest Elementary, he will have access to the data for both the CASE English Language Arts assessment and the i-Ready Diagnostic assessment as soon as it is available following the fall and winter administrations.

TEAM evaluation data for each teacher will occur throughout the 1<sup>st</sup> semester as part of the required observation timeline. All teachers in both the control and experimental group will participate in the evaluation process, and scores will be available to the researcher immediately upon completion. Unlike the two student assessments,

there will only be one data collection point for new TEAM instructional domain evaluation scores. As part of the analysis process, scores will be compared to TEAM instructional domain evaluation scores from the previous school year in two ways. First, this will be analyzed as an aggregate score for the TEAM instructional domain rubric for each teacher. Secondly, it will be analyzed on the individual selected performance indicator level. The experimental group will use the targeted performance indicator associated with their OKR that was selected during the research study. The control group will also select a performance indicator to target for improvement based on previous feedback or performance as part of the school's standard evaluation practices. The post-hoc data from the previous year will be available to the researcher immediately following IRB and district approval through personnel files kept at the school. For individual participants who do not have TEAM evaluation scores from the previous year, the researcher will use the school average from the previous year for the instruction domain.

The teacher motivation survey is administered at the end of the first semester of the Fall 2020 school year which coincides with the completion of the research study. The survey is sent out electronically using the Microsoft Forms platform to all teachers in the school with an invitation to respond. Follow-up emails are sent throughout the end of the semester to remind and encourage teachers to complete the survey. The survey window was open during the entire month of December. The Microsoft Forms platforms allow for teachers to complete the survey from any computer or mobile device and for responses to be kept anonymous. Data from the survey is available immediately after a respondent submits their responses.

## Data Analysis Procedures

Upon completion of the winter student assessment screenings and the first semester TEAM teacher evaluation cycle, the researcher will analyze results to answer each of the research questions. The researcher will utilize the *IBM SPSS* program to input data and complete descriptive analysis and inferential analysis of the data.

*RQ 1: What impact does the creating and monitoring of individual, OKR goals have on teacher performance as measured by teacher evaluation data*

Data analysis for this research question will begin by calculating an average score for each participant for the instructional domain rubric of the TEAM evaluation model. There are 12 performance indicators found within the instructional domain rubric. These 12 scores will be averaged to get a single instructional domain score for each participant. This process will be repeated with the participants' instructional domain scores from the previous school year so that there is a single score for both the former year before the introduction of the independent variable to the experimental group and one for the current year after the introduction of the independent variable. The researcher will also calculate the average instructional domain score for all teachers in the building from the previous year to generate a single school average. This average will be used for participants who do not have prior year scores. Using Microsoft Excel, post-study instructional domain averages will be subtracted from the pre-study domain averages to produce an average growth score for each participant. Data will be uploaded into the *IBM SPSS* program and descriptive statistics for the evaluation growth data for both the control group and the experimental group will be run. This data will provide the number of participants in each group as well as the mean scores and standard deviations.

After the descriptive analysis, data will be analyzed inferentially. In the analysis, we are looking to determine if there is a statistically-significant difference between our control group and experimental group on the instructional domain average growth. To test to see if this difference exists, we will use *IBM SPSS* to run an independent samples t-test. The t-test will compare the means between the two groups to determine if they are statistically significant from each other.

This analysis will be repeated using the single performance indicator score from the identified OKR aligned indicator for the experimental group and the self-selected performance indicator from the control group instead of the aggregated score for the complete domain. Using Microsoft Excel, pre-study performance indicator scores will be subtracted from the post-study performance indicator scores to produce a growth number for each participant. Data will be uploaded into the *IBM SPSS* program and descriptive statistics for the performance indicator evaluation data for both the control group and the experimental group will be run. This data will provide the number of participants in each group as well as the mean scores and standard deviations.

After the descriptive analysis, data will be analyzed inferentially. In the analysis, we are looking to determine if there is a statistically-significant difference between our control group and experimental group on the individually identified performance indicator. To test to see if this difference exists, we will use *IBM SPSS* to run an independent samples t-test. The t-test will compare the means between the two groups to determine if they are statistically significant from each other.

*RQ2: What impact does the creation and monitoring of individual, OKR goals have on student learning in the English/language arts content area as measured by student performance data?*

Data analysis for this question will begin by identifying the average percent correct score for students in 2<sup>nd</sup> – 6th grade for both the fall and winter administration of the CASE English Language Arts assessment. Using Microsoft Excel, the fall average score from each student will be subtracted from the winter average score to provide percentile growth score. Data will be uploaded into IBM SPSS and descriptive statistics for both the control and experimental groups will be run. This data will provide the number of students in each group as well as the mean scores and standard deviations.

After the descriptive analysis, data will be analyzed inferentially. In the analysis, we are looking to determine if there is a statistically-significant difference between our control group and experimental group in terms of percent correct growth on the CASE English Language Arts assessment from fall to winter. To test to see if this difference exists, we will use *IBM SPSS* to run an independent samples t-test. The t-test will compare the means between the two groups to determine if they are statistically significant from each other.

*RQ3: What impact does the creation and monitoring of individual, OKR goals have on student learning in the mathematics content area as measured by student performance data?*

Data analysis for this question will begin by identifying the scale score for students in kindergarten – 6th grade for both the fall and winter administration of the i-Ready Diagnostic assessment. Using Microsoft Excel, the fall scale score from each

student will be subtracted from the winter scale score to provide a scale score growth score. This data will be uploaded into *IBM SPSS* and descriptive statistics for both the control and experimental groups will be run. This data will provide the number of students in each group as well as the mean scores and standard deviations.

After the descriptive analysis, data will be analyzed inferentially. In the analysis, we are looking to determine if there is a statistically-significant difference between our control group and experimental group in terms of scale score growth on the i-Ready Diagnostic assessment from fall to winter. To test to see if this difference exists, we will use *IBM SPSS* to run an independent samples t-test. The t-test will compare the means between the two groups to determine if they are statistically significant from each other.

*RQ4: What impact does the creation and monitoring of individual, OKR goals have on teacher motivation as measured by perception survey data?*

Data analysis for this question will begin by recoding participant responses into a numerical value with higher numbers indicating a higher motivational perception. Table 7 below shows the recoded variables.

*Table 7*

*Numerical Recoding of Motivation Survey*

<b>Response</b>	<b>Numerical Score</b>
Strongly Agree	4
Agree	3
Disagree	2
Strongly Disagree	1

Using Microsoft Excel, the data will be organized to separate the scores of teachers in the control and experimental groups. The scores for each question will be averaged to find a mean numerical value for each question from both the control and experimental groups. Data will be uploaded into IBS SPSS and descriptive statistics for both the control and experimental groups will be run. This data will provide the number of participants in each group as well as the mean and standard deviations to each survey question.

After the descriptive analysis, data will be analyzed inferentially. In the analysis, we are looking to determine if there is a statistically-significant difference between our control group and experimental group in terms of mean numerical score for each question. To test to see if this difference exists, we will use *IBM SPSS* to run an independent samples t-test. The t-test will compare the means between the two groups to determine if they are statistically significant from each other.

### **Limitations and Delimitations**

Throughout the research study there are several limitations and delimitations that should be noted and considered in review of the research study. The study was limited to the teaching staff of an individual school. However, it is assumed that the data is drawn from a normally distributed population. In addition, the study is limited in time and scope taking place over a single semester. The researcher also serves as the principal of the school being studied; therefore, he will have a greater role in overseeing and participating in the research process. The principal will deliver the initial professional development over the OKR process, he will consult with teachers in the experimental group on their OKRs, and will provide feedback to teachers in both the control and



experimental group throughout the research. In addition, as principal, the researcher will have a supervisory role in all aspects of data collection including the student assessment data and the TEAM evaluation process. Regardless of this joint role as principal and researcher, all teachers in the building would be participating in the student assessment and TEAM evaluation process even if the research study were not taking place. No additional data collection is being required based on the research study. Participants randomly selected to join the study will be informed that participation is voluntary and that they can withdraw from the study at any time.

Creswell and Creswell (2018) emphasize the importance of identifying potential threats to validity and planning actions in response to the threat. In describing the history threat of internal validity, Creswell and Creswell (2018) state, “events can occur that unduly influence the outcome beyond the experimental treatment” (p. 170). Both the state of Tennessee where the study is being completed and the nation are currently experiencing the COVID-19 pandemic. In response to this, school systems are planning to make modifications to the school day to ensure public safety. Although, the potential of interrupted schooling could impact student achievement, the external events would be experienced by both the control and experimental groups therefore minimizing threats to validity it may cause.

### **Summary**

Throughout chapter three, the design for a quasi-experimental study has been outlined that will determine if there is a statistically-significant difference in student and teacher performance outcomes when teachers implement the OKR system as a method of improvement. In addition to defining the research setting and methodology, a clear

rationale for the importance of both the study's topic and research design were given. Data sources, data collection procedures, and data analysis procedures were specified as were limitations, delimitations, and potential threats of validity. By implementing this study, educators will be provided additional insight on the potential benefits of implementing improvement methods that may not be specifically designed for the school environment. The table below summarizes the chapter by clearly identifying the research questions, the data sources used, and the data analysis procedures.

Table 8

*Research Question and Data Summary Table*

<b>Research Question</b>	<b>Data Source</b>	<b>Data Analysis Procedure</b>
RQ1: What impact does the creating and monitoring of individual, OKR goals have on teacher performance as measured by teacher evaluation data?	TEAM evaluation scores from instructional domain rubric.	Descriptive Statistics Independent samples t-test
RQ2: What impact does the creation and monitoring of individual, OKR goals have on student learning in the English/language arts content area as measured by student performance data?	CASE English Language Arts benchmark assessment data from the fall and winter administration	Descriptive Statistics Independent samples t-test
RQ3: What impact does the creation and monitoring of individual, OKR goals have on student learning in the mathematics content area as measured by student performance data?	i-Ready Diagnostic assessment data from the fall and winter administration	Descriptive Statistics Independent samples t-test
RQ4: What impact does the creation and monitoring of individual, OKR goals have on teacher motivation as measured by perception survey data?	Creation and monitoring of individual, OKR goals	Responses indicated on the motivation survey.

## **CHAPTER IV**

### **FINDINGS**

#### **Introduction**

Examining the impact of goal setting through the Objective and Key Result (OKR) process was the primary purpose in this quasi-experimental study. Specifically, the researcher intended to identify whether or not there was a statistically significant difference in student and teacher outcomes when the OKR process was implemented with a random sampling of classroom teachers in a suburban elementary school. This chapter provides a descriptive understanding of both the control and experimental groups. In addition, analyzed results from the data sources collected in the research are reported.

#### **Research Questions**

In this study, the following research questions are answered:

1. What impact does the creating and monitoring of individual, OKR goals have on teacher performance as measured by teacher evaluation data?
2. What impact does the creation and monitoring of individual, OKR goals have on student learning in the English/language arts content area as measured by student performance data?
3. What impact does the creation and monitoring of individual, OKR goals have on student learning in the mathematics content area as measured by student performance data?
4. What impact does the creation and monitoring of individual, OKR goals have on teacher motivation as measured by perception survey data?

### **School and Research Group Profiles**

Southwest Elementary School is a kindergarten through sixth grade, suburban public school located in the state of Tennessee. Southwest Elementary was in the second year of its existence when the study was completed. The timeline of the study also coincided with the school district implementing new procedures in response to the COVID-19 pandemic that was occurring across the United States. The adjustments that were made to the daily operation of the school impacted normal schooling procedures. These changes resulted in some students learning in a virtual environment while others were learning in person. In addition, teachers' instructional methods were a mix of all in-person, all virtual, or teaching in hybrid classrooms that included both in-person and virtual students. The variety of learning models also resulted in student assessments being delivered in a variety of formats that impacted the consistency of the testing environment.

As one of thirteen schools in the Central Schools District, Southwest has a diverse school population of over 900 students during their second year. Table 9 below outlines important demographic information for Southwest compared to district and state data as identified at the end of the school's inaugural year.

Table 9

*Percentage of Historically Underserved Students*

	<b>Southwest Elementary</b>	<b>Central Schools District</b>	<b>State</b>
<b>Black, Hispanic, and Native American</b>	33.9%	45.2%	35.2%
<b>Economically Disadvantaged</b>	21.3%	32%	34.9%

The experimental group in the study was pulled from a random sampling of the population of 45 kindergarten through 6<sup>th</sup> grade teachers at Southwest Elementary School. The study was limited to only full-time classroom teachers. Four teachers were excluded from the study prior to participant selection due to known medical reasons that would prevent them from being the primary instructor of their classrooms for a significant period of time thus preventing them from delivering a large portion of the instruction during the study period. The students in these four classrooms were also excluded from the study. Table 10 below details important information on the make-up of the 41 teachers who made up the research population.

Table 10

*Characteristics of the Teacher Population*

<b>Average Years of Experience</b>	<b>Range of Experience</b>	<b>Number with 0 years' Experience</b>	<b>Number with less than 5 Years' Experience</b>	<b>Number New to Southwest</b>
7.2	0 years – 19 years	5	19	15

From this population of teachers, 21 teachers were selected through random sampling to participate in the experimental group. The characteristics of the experimental group are detailed below in Table 11.

Table 11

*Characteristics of the Teacher in the Experimental Group*

<b>Average Years of Experience</b>	<b>Range of Experience</b>	<b>Number with 0 years' Experience</b>	<b>Number with less than 5 Years' Experience</b>	<b>Number New to Southwest</b>
8.46	0 years – 19 years	3	7	10

Table 12 details the grades taught by the experimental group.

Table 12

*Grade Make-Up of Experimental Group*

Kindergarten	1 <sup>st</sup> Grade	2 <sup>nd</sup> Grade	3 <sup>rd</sup> Grade	4 <sup>th</sup> Grade	5 <sup>th</sup> Grade	6 <sup>th</sup> Grade
4	2	5	3	2	4	1

Once the 21 experimental teachers were removed from the population, 20 teachers remained and served as the control group for the study. The characteristics of the control group are detailed below in table 13.

Table 13

*Characteristics of the Teacher in the Control Group*

Average Years of Experience	Range of Experience	Number with 0 years' Experience	Number with less than 5 Years' Experience	Number New to Southwest
6.2 years	0 years – 19 years	2	12	5

Table 14 details the grade make-up of the group by listing the number of teachers from each grade level who participated in the control group.

Table 14

*Grade Make-Up of Control Group*

Kindergarten	1 <sup>st</sup> Grade	2 <sup>nd</sup> Grade	3 <sup>rd</sup> Grade	4 <sup>th</sup> Grade	5 <sup>th</sup> Grade	6 <sup>th</sup> Grade
3	6	2	5	3	0	1



The charts above demonstrate that the control group did have, on average, slightly lower average experience than those in the experimental group by 2.26 years. Although, the experimental group does contain teachers with a higher experience level, it also contains notably more teachers, 10, who are in the first year of teaching at Southwest Elementary School. This is double the number of new to Southwest teachers found in the control group.

### **Research Question 1**

Research question 1 sought evidence to answer the following question: What impact does the creating and monitoring of individual, OKR goals have on teacher performance as measured by teacher evaluation data?

Evaluation data from the state-approved TEAM evaluation system was used as the data source for this question. The evaluation data collected during the research period was compiled and analyzed against the evaluation data from the previous school year in two ways. First, scores were analyzed as an aggregated score for the instructional domain rubric as a whole for each teacher. This aggregate score is reported as the average of the 12 performance indicators that make up a single score for the instructional domain rubric ranging from 1 to 5. Secondly, scores are analyzed on one individually chosen performance indicator that was associated with the experimental groups' OKR goal or was chosen as a focus area by individuals in the control group. Similarly, to the domain score, the performance indicator score is also reported on a 1 to 5 scale.

By analyzing scores in two ways, the researcher intended to examine whether or not there was a statistically significant difference in teacher performance as a whole as

measured by the TEAM evaluation rubric and/or whether that difference was evident in the area that tied closest to the participants' OKR.

*Analysis Using TEAM Instructional Domain Average Score*

Table 15 below shows the descriptive statistics of the population for the aggregate instruction domain rubric while Table 16 shows the same descriptive statistics broken down by control and experimental group.

Table 15

*Descriptive Statistics for the Population's Instruction Domain Scores*

	Minimum	Maximum	Mean	Standard Deviation
<b>2019-2020 Domain Average</b>	2.67	4.83	3.5562	.70548
<b>2020-2021 Domain Average</b>	2.21	4.75	3.6686	.67327
<b>Domain Growth Average</b>	-.83	1.25	.1124	.45626

Table 16

*Descriptive Statistics for the Control and Experimental Groups Instruction Domain**Scores*

		<b>N-Count</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Standard Error Mean</b>
<b>2019-2020 Domain Average</b>	Control	20	3.3236	.66528	.14876
	Experimental	21	3.7792	.71698	.15646
<b>2020-2021 Domain Average</b>	Control	20	3.5750	.66757	.14927
	Experimental	21	3.7341	.71352	.15570
<b>Domain Growth Average</b>	Control	20	.2514	.39738	.08886
	Experimental	21	-0.0450	.48017	.10478

The data presented above indicates that the mean score for the entire population did increase by .1124 points from the first year to the second year. When this information is separated between the control and experimental groups, we can see that the growth was made by the control group. It is important to recognize that although the experimental group did not show growth from the 2019-2020 school year to the 2020-2021 school year, they did start off with a higher average mean score by .45 points. After the growth seen by the control group, their mean score was still .15 points less than the experimental group.

To determine if a statistically-significant difference between the control and experimental group exists, *IBM SPSS* was used to run an independent samples t-test to compare the mean growth between the two groups. Table 17 below shows the results of the independent samples t-test.

Table 17

*Independent Samples t-test for Instructional Domain Rubric*

		Levene's Test for Equality of Variance		t-test for Equality of Means				95% Confidence Interval of the Difference		
		F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
	Equal variances assumed	1.740	.195	2.148	39	.038	.29644	.13803	.01725	.57563
Domain Growth	Equal variances not assumed			2.158	38.274	.037	.29644	.13739	.10838	.57449

The results of this study found a statistically significant difference between the control and the experimental group in their growth on the TEAM instructional domain aggregate score from the 2019-2020 school year to the 2020-2021 school year,  $t(39) = 2.148$ ,  $p=.038$ , so we reject the null hypothesis that there is no difference. However, this data indicates that the control group actually had the higher mean TEAM instructional domain score than the experimental group. Although this data does present a statistical difference, it is not supportive of the hypothesis that the use of OKRs would improve teacher instructional performance as measured by TEAM.

*Analysis Using Participants Selected Focus Performance Indicator*

Table 18 below shows the descriptive statistics of the population for the individual performance indicator found within the instructional domain rubric that was specifically identified as a focus area by the participant. It should be noted that 12 participants in the control group failed to identify a focus area which impacted the N-count.

Table 18

*Descriptive Statistics for the Population's Identified Performance Indicator Focus Area**Score*

	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Standard Deviation</b>
<b>2019-2020 Domain Average</b>	2.00	5.00	3.5556	.89994
<b>2020-2021 Domain Average</b>	2.00	5.00	3.6333	.71840
<b>Domain Growth Average</b>	-1.50	2.00	.0778	.85077

As represented in the table above, the mean score for the individual focus indicator did improve from the 2019-2020 school year to the 2020-2021 school year.

Table 19 shows the same descriptive statistics broken down by control and experimental group.

Table 19

*Descriptive Statistics for the Control and Experimental Groups Identified Performance**Indicator Focus Area Score*

		<b>N-Count</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Standard Error Mean</b>
<b>2019-2020 Domain Average</b>	Control	8	3.0069	.87612	.30975
	Experimental	21	3.7910	.84360	.18409
<b>2020-2021 Domain Average</b>	Control	8	3.6250	.79057	.27951
	Experimental	21	3.6190	.72292	.15775
<b>Domain Growth Average</b>	Control	8	.6181	.81997	.28990
	Experimental	21	-0.1720	.76775	.16754

Similar to what was found when looking at the descriptive data of the instructional domain score, the control group was identified as having more growth than the experimental group in their individual focus performance indicator. In contrast to the analysis of the instructional domain average score, the growth of the control group took them from an average mean score below the experimental group in 2019-2020 to an average mean score above the experimental group in 2020-2021. However, the lower N-count of control group should be considered when reviewing these results.

To determine if a statistically-significant difference between the control and experimental group exists, *IBM SPSS* was used to run an independent samples t-test to compare the mean growth on the individually chosen performance indicator between the two groups. Table 20 below shows the results of the independent samples t-test.

Table 20

*Independent Samples t-test for Identified Performance Indicator Focus Area Score*

		Levene's Test for Equality of Variance		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Domain Growth	Equal variances assumed	.003	.956	2.433	27	.022	.79001	.32474	.12369	1.45633
	Equal variances not assumed			2.359	11.988	.036	.79001	.33483	.06040	1.51962

The results of this study found a statistically significant difference between the control and the experimental group in their growth on the identified performance indicator focus area score from the TEAM instructional domain rubric from the 2019-2020 school year to the 2020-2021 school year,  $t(27) = 2.433$ ,  $p = .022$ , so we reject the null hypothesis that there is no difference. However, this data indicates that the control group has a higher mean TEAM performance indicator focus area growth score than the experimental group. Although this data does present a statistical difference, it is not supportive of the hypothesis that the use of OKRs would improve teacher instructional performance as measured by TEAM.

## Research Question 2

Research question 2 sought evidence to answer the following question:

What impact does the creation and monitoring of individual, OKR goals have on student learning in the English/language arts content area as measured by student performance data?

The Case English Language Arts assessments were used as the data source to answer this research question. The CASE assessment was administered to 2<sup>nd</sup> through 6<sup>th</sup> grade students at the beginning of the school year and a second equivalent form was administered at the mid-year point of the school year. The researcher examined the average percent correct score from the first and second assessments to identify differences between the groups.

Table 21 below shows the descriptive statistics of the population for the first (T1) and second (T2) assessment while Table 22 shows the same descriptive statistics broken down by control and experimental group.

Table 21

*Descriptive Statistics for the Population's Total T1 and Total T2 Case Scores*

	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Standard Deviation</b>
<b>Total Score T1</b>	490	5.7	100	49.418	20.1374
<b>Total Score T2</b>	490	2.5	96.7	50.826	19.7414
<b>Total Score +/-</b>	490	-45.4	55.0	1.409	13.7222



Table 22

*Descriptive Statistics for the Control and Experimental Groups Total T1 and Total T2**Case Scores*

		<b>N-Count</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Standard Error Mean</b>
<b>Total Score T1</b>	Control	182	45.718	19.0286	1.4105
	Experimental	292	51.114	20.1425	1.1788
<b>Total Score T2</b>	Control	182	48.929	19.1098	1.4165
	Experimental	292	51.950	20.0793	1.1751
<b>Total Score +/-</b>	Control	182	3.212	12.2764	.9100
	Experimental	292	.846	14.2620	.8346

The data presented above indicates that the mean score for the entire population did increase by 1.409 points from the T1 to T2. When this information is separated between the control and experimental groups, we can see that a larger growth was made by the control group. It is important to recognize that although the experimental group did not show as much growth from T1 to T2, they did start off with a higher average mean score by 5.396 points. After the growth seen by the control group, their mean score was still 3.021 points less than the experimental group's T2 score and 2.185 less than the experimental group's T1 score.

To determine if a statistically-significant difference between the control and experimental group exists, *IBM SPSS* was used to run an independent samples t-test to compare the mean growth between the two groups. Table 23 below shows the results of the independent samples t-test.

Table 23

*Independent Samples t-test for Case Assessment Total Score +/-*

		Levene's Test for Equality of Variance		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Percent Correct Growth	Equal variances assumed	4.315	.038	1.851	472	.065	2.3660	1.2783	-1.458	4.8778
	Equal variances not assumed			1.916	426.070	.056	2.3660	1.2348	-.0610	4.7930

The results of this study did not find a statistically significant difference between the control and the experimental group in their growth on the Case assessment total percent correct from the first to the second test administration,  $t(426.070) = 1.916$ ,  $p=.56$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group, the lack of statistical significance suggests that there is no real difference.

### **Research Question 3**

Research question 3 sought evidence to answer the following question:

What impact does the creation and monitoring of individual, OKR goals have on student learning in the mathematics content area as measured by student performance data?

The i-Ready Diagnostic math assessments were used as the data source to answer this research question. This assessment was administered to Kindergarten through 6<sup>th</sup> grade students at the beginning of the school year and a second equivalent form was administered at the mid-year point of the school year. The researcher examined the

average scale score from the first and second assessment to identify differences between the groups.

Table 24 below shows the descriptive statistics of the population first (T1) and second (T2) assessment while table 25 shows the same descriptive statistics broken down by control and experimental group.

Table 24

*Descriptive Statistics for the Population's Total T1 and Total T2 i-Ready Scores*

	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Standard Deviation</b>
<b>T1 Scale Score</b>	876	277	539	406.56	48.984
<b>T2 Scale Score</b>	876	287	548	422.40	45.242
<b>Total Scale Score +/-</b>	876	-118	120	15.84	20.498

Table 25

*Descriptive Statistics for the Control and Experimental Groups Total T1 and Total T2 i-Ready Scores*

		<b>N-Count</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Standard Error Mean</b>
<b>T1 Scale Score</b>	Control	358	398.72	42.158	2.228
	Experimental	443	415.90	51.779	2.460
<b>T2 Scale Score</b>	Control	358	415.42	39.980	2.113
	Experimental	443	430.79	47.453	2.255
<b>Total Scale Score +/-</b>	Control	358	16.70	23.749	1.255
	Experimental	443	14.89	18.104	.850

The data presented above indicates that the mean score for the entire population did increase by 15.84 scale score points from the T1 to T2. When this information is separated between the control and experimental groups, we can see that a larger growth was made by the control group with a 16.70 average scale score increase compared to a 14.89 average scale score increase for the experimental group. It is important to recognize that although the experimental group did not show as much growth from T1 to T2, they did start off with a higher average scale score by 17.18 scale score points. After the growth seen by the control group, their mean score was still 15.37 scale score points less than the experimental group's T2 score and only .48 scale points higher than the experimental group's initial T1 score.

To determine if a statistically-significant difference between the control and experimental group exists, *IBM SPSS* was used to run an independent samples t-test to compare the mean growth between the two groups. Table 26 below shows the results of the independent samples t-test.

Table 26

*Independent Samples t-test for i-Ready Assessment Total Score +/-*

		Levene's Test for Equality of Variance		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Scale Score Growth	Equal variances assumed	5.401	.020	1.225	799	.221	1.812	1.479	-1.092	4.716
	Equal variances not assumed			1.191	654.464	.234	1.812	1.522	-1.176	4.800

The results of this study did not find a statistically significant difference between the control and the experimental group in their growth on the i-Ready Diagnostic math assessment average scale score from the first to the second test administration,  $t(654.464) = 1.191$ ,  $p=.234$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group, the lack of statistical significance suggests that there is no real difference.

#### **Research Question 4**

Research question 4 sought evidence to answer the following question:  
What impact does the creation and monitoring of individual, OKR goals have on teacher motivation as measured by perception survey data?

An eight-question survey was sent out to teachers in both the control and experimental group. The survey questions were taken directly from the Tennessee Educator Survey which is administered to Tennessee teachers each year. Table 27 below displays the eight questions asked on the survey.

Table 27

*Motivation Survey Questions*


---

<b>Q1</b>	I am generally satisfied being a teacher in this school.
<b>Q2</b>	Our school staff is a learning community in which ideas and suggestions for improvement are encouraged.
<b>Q3</b>	Our staff feels comfortable raising issues and concerns that are important to them with school leaders.
<b>Q4</b>	My administration regularly gives feedback on my instruction.
<b>Q5</b>	My administration knows my instructional strengths and areas of growth.
<b>Q6</b>	Teachers at my school feel responsible to help each other do their best.
<b>Q7</b>	Teachers at my school take responsibility for improving their school.
<b>Q8</b>	In general, the teacher evaluation process used in my school has led to improvements in my teaching.

---

Forty participants responded to the questions with one of the following four options: strongly agree, agree, disagree, strongly disagree. These responses were recoded into numerical values as described in table 28 below.

Table 28

*Numerical Recoding of Motivation Survey*

<b>Response</b>	<b>Numerical Score</b>
Strongly Agree	4
Agree	3
Disagree	2
Strongly Disagree	1

The researcher examined the average scale score each of the eight questions to identify differences between the groups.

Table 29 below shows the descriptive statistics of the population's survey results.

Table 29

*Descriptive Statistics for the Population's Survey Results*

	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Standard Deviation</b>
<b>Q1</b>	40	3	4	3.82	.385
<b>Q2</b>	40	3	4	3.78	.423
<b>Q3</b>	40	2	4	3.55	.597
<b>Q4</b>	40	2	4	3.53	.599
<b>Q5</b>	40	3	4	3.68	.474
<b>Q6</b>	40	2	4	3.62	.540
<b>Q7</b>	40	3	4	3.70	.464
<b>Q8</b>	40	2	4	3.45	.597

The data presented above indicates that range of mean scores is 3.45 – 3.82. Question 1, *I am generally satisfied being a teacher in this school*, showed the highest mean score. Question 8, *In general, the teacher evaluation process used in my school has led to improvements in my teaching*, had the lowest average score. It is important to note that the difference between the question with the highest average score and the question with the lowest average score was 0.37. Table 30 below shows the same descriptive statistics as above but broken down by control and experimental group.



Table 30

*Descriptive Statistics for the Control and Experimental Groups Motivation Survey**Results*

	<b>N-Count</b>	<b>N-Count</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Standard Error Mean</b>
<b>Q1</b>	Control	24	3.88	.338	.069
	Experimental	16	3.75	.447	.112
<b>Q2</b>	Control	24	3.88	.338	.069
	Experimental	16	3.63	.500	.125
<b>Q3</b>	Control	24	3.58	.584	.119
	Experimental	16	3.50	.632	.158
<b>Q4</b>	Control	24	3.63	.576	.118
	Experimental	16	3.38	.619	.115
<b>Q5</b>	Control	24	3.75	.442	.090
	Experimental	16	3.56	.512	.128
<b>Q6</b>	Control	24	3.75	.532	.109
	Experimental	16	3.44	.512	.128
<b>Q7</b>	Control	24	3.79	.415	.085
	Experimental	16	3.56	.512	.128
<b>Q8</b>	Control	24	3.50	.590	.120
	Experimental	16	3.38	.619	.155

When this data is separated between the control and experimental groups, we can see that the control group had a higher average mean score for each of the eight

questions. In addition, the control group had a larger population with 24 participants which was 8 more than participated in the experimental group.

To determine if a statistically-significant difference between the control and experimental group exists, *IBM SPSS* was used to run an independent samples t-test to compare the mean score between the two groups for each question. Table 31 through Table 38 below shows the results of the independent samples t-test.

Table 31

*Independent Samples t-test for Motivation Survey Results Question 1*

	Levene's Test for Equality of Variances		t-test Equality of Means		95% Confidence Interval of the Difference				
	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	4.00	.053	1.007	38	.320	.125	.124	-.126	.376
<b>Q1</b> Equal variances not assumed			.952	26.119	.350	.125	.131	-.145	.395

The results of this study did not find a statistically significant difference between the control and the experimental group in their responses to question 1 on the motivation survey,  $t(38) = 1.007$ ,  $p = .320$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group for question 1, the lack of statistical significance suggests that there is no real difference.

Table 32

*Independent Samples t-test for Motivation Survey Results Question 2*

	Levene's Test for Equality of Variances		t-test Equality of Means			95% Confidence Interval of the Difference			
	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	13.326	.001	1.891	38	.066	.250	.132	-.018	.518
<b>Q2</b> Equal variances not assumed			1.751	24.066	.093	.250	.143	-.045	.545

The results of this study did not find a statistically significant difference between the control and the experimental group in their responses to question 2 on the motivation survey,  $t(24.066) = 1.751$ ,  $p=.093$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group for question 2, the lack of statistical significance suggests that there is no real difference.

Table 33

*Independent Samples t-test for Motivation Survey Results Question 3*

	Levene's Test for Equality of Variances		t-test Equality of Means		95% Confidence Interval of the Difference				
	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.280	.600	.428	38	.671	.083	.195	-.311	.478
<b>Q3</b> Equal variances not assumed			.421	30.461	.677	.083	.198	-.321	.487

The results of this study did not find a statistically significant difference between the control and the experimental group in their responses to question 3 on the motivation survey,  $t(38) = .428$ ,  $p = .671$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group for question 3, the lack of statistical significance suggests that there is no real difference.

Table 34

*Independent Samples t-test for Motivation Survey Results Question 4*

	Levene's Test for Equality of Variances		t-test Equality of Means			95% Confidence Interval of the Difference			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.310	.581	1.306	38	.200	.250	.191	-.138	.638
<b>Q4</b> Equal variances not assumed			1.286	30.640	.208	.250	.194	-.147	.647

The results of this study did not find a statistically significant difference between the control and the experimental group in their responses to question 4 on the motivation survey,  $t(38) = 1.306$ ,  $p = .200$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group for question 4, the lack of statistical significance suggests that there is no real difference.

Table 35

*Independent Samples t-test for Motivation Survey Results Question 5*

	Levene's Test for Equality of Variances		t-test Equality of Means			95% Confidence Interval of the Difference				
	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
<b>Q5</b>	Equal variances assumed	4.222	.047	1.233	38	.225	.188	.152	-.120	.495
	Equal variances not assumed			1.196	28.949	.241	.188	.157	-.133	.508

The results of this study did not find a statistically significant difference between the control and the experimental group in their responses to question 5 on the motivation survey,  $t(28.949) = 1.196$ ,  $p=.241$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group for question 5, the lack of statistical significance suggests that there is no real difference.

Table 36

*Independent Samples t-test for Motivation Survey Results Question 6*

	Levene's Test for Equality of Variances		t-test Equality of Means		95% Confidence Interval of the Difference				
	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	1.209	.278	1.847	38	.072	.313	.169	-0.30	.655
<b>Q6</b> Equal variances not assumed			1.862	33.129	.072	.313	.168	-0.29	.654

The results of this study did not find a statistically significant difference between the control and the experimental group in their responses to question 6 on the motivation survey,  $t(38) = 1.847$ ,  $p = .072$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group for question 6, the lack of statistical significance suggests that there is no real difference.

Table 37

*Independent Samples t-test for Motivation Survey Results Question 7*

	Levene's Test for Equality of Variances		t-test Equality of Means		95% Confidence Interval of the Difference				
	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	6.825	.013	1.558	38	.128	.229	.147	-.069	.527
<b>Q7</b> Equal variances not assumed			1.492	27.546	.147	.229	.154	-.086	.544

The results of this study did not find a statistically significant difference between the control and the experimental group in their responses to question 7 on the motivation survey,  $t(27.546) = 1.492$ ,  $p=.147$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group for question 7, the lack of statistical significance suggests that there is no real difference.



Table 38

*Independent Samples t-test for Motivation Survey Results Question 8*

	Levene's Test for Equality of Variances		t-test Equality of Means		95% Confidence Interval of the Difference				
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.005	.943	.644	38	.524	.125	.194	-.268	.518
<b>Q8</b> Equal variances not assumed			.637	31.192	.528	.125	.196	-.275	.525

The results of this study did not find a statistically significant difference between the control and the experimental group in their responses to question 8 on the motivation survey,  $t(38) = .644$ ,  $p = .524$ , so we fail to reject the null hypothesis that there is no difference. Although the descriptive data reflected a slightly higher mean percent correct growth score for the control group for question 8, the lack of statistical significance suggests that there is no real difference.

### **Conclusion**

Chapter 4 contains data specifically designed to measure the difference between two groups of classroom teachers when one of those groups is exposed to the independent variable of setting OKR goals. Through the analysis of this quantitative data, it has shown that the OKR process did not have a significant impact on student performance outcomes or teacher motivation. However, it did have a significant impact on teacher performance as measured by TEAM scores. Unfortunately, that was significance in the negative

direction. As such, the findings of this study suggest that schools and districts may not see immediate benefit from changing their goal-setting approach through adaptation of the OKR model. The following chapter will provide additional context around implications, conclusions and interpretations of the data that may shed further information on why this analysis did not show the results that are implied in the literature review.

## CHAPTER V

### CONCLUSIONS

#### Introduction

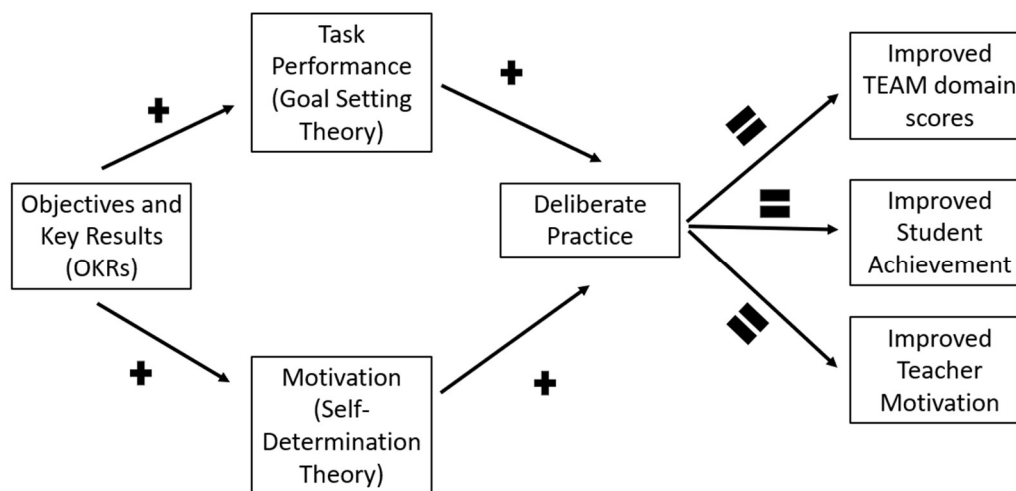
Throughout the past four chapters this work has outlined multiple discussion points that seek to drive an understanding towards a new method of improving the performance of teachers and the educational outcomes of students. This began with the understanding that school improvement and school reform initiatives have continuously been one of the most prevalent and controversial topics in the field of education. Although there has been a lack of consensus around improvements, research (Hattie, 2012, Darling-Hammond & Goldberg, 2001) has pointed to the key role the teacher has in influencing learning outcomes of students. However, the number of reform initiatives at play in the field of education has the potential of leading teachers down multiple improvement paths that may result in a scattered approach to their improvement process.

Knowing this, the purpose of this study was to focus on the impact of a singular improvement method that provided a clear framework to educators while also allowing them autonomy to meet the contextual needs of their situation. Doerr's (2018) Objective and Key Result (OKR) model has been used in various businesses, and in his book *Measure What Matters*, he details how his process has been attributed to the highly successful growth and performance of companies. In this study, the OKR process was implemented in a quasi-experimental design to test if this practice might also translate to success in the K12 environment and the school improvement process. By analyzing qualitative data, the researcher sought to identify whether or not there was a statistically significant difference in student and teacher outcomes when classroom teachers utilize

this same goal setting method by setting clear objectives and actionable key results as described by Doerr (2018).

Two prominent, research-based theories grounded this work and provided the context that would seemingly support the hypothesis that implementing OKRs would in fact lead to improved outcomes. The first theory is Locke and Latham's (1990) goal-setting theory which clearly connects an individual's enhanced performance to the practice of setting goals that are specific, high, and attainable over performances that are guided without clear goals or goals that are too easy to maintain. Ryan and Deci's (2000) Self-Determination Theory (SDT) further supported the work of OKRs by outlining how the process should increase the intrinsic motivation of teachers to reach their goals while simultaneously helping support a culture built on mutual trust and respect. Ultimately, these two theories reinforced the ideas that setting OKRs would help teachers take ownership and be more aware of improvement goals while also increasing their motivation to achieve the goals. As shown in chapter one, the logic model below outlines the hypothesis of the quasi-experimental study.

Figure 4

*Logic Model*

The context of this research was confined to the classroom teachers in a single elementary school over the first semester of a school year. The random selection of teachers was used to populate both the control group and experimental group who would participate in the OKR process.

### Summary of Results

The review of research and learning theories previously discussed clearly supported the use of OKRs as a potentially significant improvement model for education. In light of both the theoretical and empirical evidence supporting the OKR process, four specific research questions were examined in the school setting to evaluate the impact the process would have on teacher performance, student achievement in English/language art and math, and teacher motivation. However, in this study, the results did not show that outcomes were improved by the implementation of the goal setting model.

The first research question looked specifically at the impact of OKRs on teacher performance by using the teachers' required evaluation data as the measurement source. Analyzing the data in two ways, the researcher sought to see if there was a change in the teachers' aggregate evaluation score and/or with an individual focus area score that aligned to the teachers' goals. The analysis of both the aggregate evaluation growth score and the focus area growth score showed that there was a statistically significant difference between the control and experimental groups. However, this difference did not support the use of OKRs as a method of improving a teacher's instructional performance because the control group's growth score averages were higher than the experimental group who implemented OKRs.

Research questions two and three focused on student outcomes by analyzing their achievement growth on district reading and math benchmark assessments that were given to all students in the district. The hypothesis predicted that by using the OKR process the teacher would positively impact the students learning outcomes and that students taught by teachers who implemented OKRs would have a higher average growth score from the first to second assessment. The analysis of both the reading and math data, however, failed to show a statistically significant difference between the control and experimental groups growth scores. In addition, the control group actually showed a slightly higher average mean growth score in both subject areas.

The final research question looked to analyze if a teacher's motivation was impacted by implementing this goal-setting process. By using an eight-question perception survey, the hypothesis was that teachers implementing the OKR model would show increased levels of motivation over the control group. Similarly, to the previous

research questions on student performance, there was not a statistically significant difference between the two groups on any of the eight questions. Also similar to the student outcome questions, the control group had a higher average mean score on each of the eight questions suggesting they were more motivated than the group implementing OKRs.

These results suggest that it may not be worth the effort to implement a full shift from traditional goal-setting approaches for teacher development to this specific OKR process. In addition, although there was not any statistically significant difference, the experimental group actually under performed on the descriptive data as compared to those who did not use the process so it is not just a matter of not enough statistical power.

## **Discussion**

Lencioni (2002) articulates that an “unrelenting focus on specific objectives and clearly defined outcomes is a requirement for any team that judges itself on performance” (p. 216), and Doerr (2018) contends that OKRs are a “launchpad” (p. 245) that he hopes will transform every walk of life including education. However, the results of this study seem to indicate that a focus on clear objectives and key results may not produce the outcomes desired. Therefore, it is important to recognize where this study’s specific findings fit within the larger pool of research on the topic, and how they may be impacted in the educational environment. Specifically, the next sections further discuss the results and context around the area of teacher improvement, student performance improvement, and teacher motivation.

### *Teacher Improvement*

Locke and Latham's (2002) goal-setting theory indicates that the setting of goals is important in the work of improving outcomes. Put more simply, they emphasize that when people are only asked to do their best, they do not. Instead, their work contends that setting specific, difficult goals produce an "energizing function" (p. 706) which leads to higher levels of effort and performance. When considering this mismatch between the literature and this study's results in regard to the impact on teacher evaluation, the specific characteristics of the two groups provide additional information that shed some light on the discrepancy.

The experimental group who participated in the OKR process contained twice the number of teachers who were serving in their first year at Southwest Elementary as the control group did. This means that their 2019-2020 evaluation scores were not completed by the current school administration at a higher rate than the control group who had fewer teachers in their first year at Southwest. Having a different evaluator will lend itself to a variation in the scores even though all TEAM evaluators must be certified through a state designed norming process. Weisburg et al. (2009) and Kraft and Gimour (2017) support this idea by discussing that new evaluation methods such as TEAM have not led to greater differentiation among performance ratings teachers are assigned. Further highlighting the discrepancy of scoring, a Tennessee Department of Education report (year p. 30 of lit review) highlights a 93 percent misalignment between teacher evaluation scores and student growth data for students whose growth data places teachers in the lowest effectiveness level of 1 and a similar misalignment for 69 percent of teachers whose students' scores placed them in the second lowest level of effectiveness.



This misalignment is illustrated in the discrepancy between the 2019-2020 average scores of the two groups which served as the baseline. In 2019-2020, the experimental group, who had the notably higher number of new to Southwest teachers, had an average instructional domain scores that was .45 points higher than the control group. The control group's growth of .2514 points in 2020-2021 still put their average score .20 points under the experimental groups 2019-2020 baseline score. Considering the scale of possible scores on the TEAM evaluation is from a 1 to a 5, the experimental group's 2019-2020 base score of 3.7792 base score and 2020-2021 score of 3.7341 represents a high average for both years which could factor in on the lack of growth for this group. This is further supported by the fact that even with the high growth shown by the control group, they still did not surpass the average score for the experimental group.

This large discrepancy in average TEAM scores is also found in the second analysis which centered in on one focus area. In this data set there was a .79 difference between the control group's 2019-2020 baseline average score of 3.0069 and experimental group's 2019-2020 average score of 3.7910. Similar to what was shown when looking at the aggregate domain score, the control group's 2020-2021 higher score still fell below the experimental group's 2019-2020 baseline score by .16 points. This data indicates that future studies involving the impact on TEAM evaluation scores may benefit from purposefully examining control and experimental groups that are not randomized but are instead stratified with equal numbers of participants representing each score 1 to 5.

A second factor that could provide additional information on the disconnect between the literature and results of this study is found in how the teachers in the study

embraced the concept of sharing their goals publicly. Doerr (2018) emphasizes the importance of ensuring goals are clear not just to the person who has the goals but to others in the organization as well. He states, “When goals are public and visible to all, a ‘team of teams’ can attack trouble spots wherever they surface” (Doerr, 2018, p. 89). Additional goal setting research (Camp, 2017, Locke & Latham, 2002, MacLeod, 2012) also emphasizes that public goal setting is a key factor in increasing commitment to the goal which leads to stronger performance. In this study, participants shared their OKRs with their supervisor and were encouraged to share them with others in the study.

However, there was not a clear protocol that required participants to share beyond the supervisor nor was there a measurement tool to identify to what extent if any that they were shared publicly. A larger emphasis on sharing OKRs with those on their grade level team regardless of whether the rest of the team was participating in the study could have helped garner the group’s support in attacking the goal as described by Doerr (2018).

As mentioned in Chapter two, the PLC SMART Goal process is a current goal-setting protocol that is prevalent in many schools as the popularity of the PLC process continues to gain traction in education. Where OKRs emphasize public goals, DuFour et al. (2016) emphasizes that PLC teams should be collaboratively writing collective goals to address the needs of their students. By writing goals collaboratively, PLC teams are by default making the goals public to those they work closest with and are therefore able to collaborate on strategies to reach the goal. With this in mind, future research in this area could consider blending the OKR and PLC SMART goal process to better harness the power of public goal sharing. This could include PLC teams collaboratively creating team objectives with individual team members having varied key results to meet the team

objective. Another variation would be to ensure all members of the PLC team are involved in the OKR process so that they could publicly and consistently review each person's individual OKR and progress in meeting the goals. Ultimately, the public nature of goals is about transparency and clarity to ensure all members of the team know what is being worked on. In this study, it is possible that because the PLC teams had both control group participants and experimental group participants the public goal sharing factor could have been impacted.

### *Student Performance*

In education, improvement is not only measured by the actions or evaluation of the teacher's performance, but it is also measured by the growth and achievement of the students who sit in tutelage of the instructor. Therefore, the goal of any improvement model must result in increased outcomes for students. As DuFour et al. (2016) emphasize, "efforts must ultimately be assessed on the basis of results rather than intentions" (p.12).

As previously discussed, the results of this study did not lead to a statistically significant increase in academic growth of students as hypothesized. However, much like the area of teacher performance, a further look at the characteristics of the students show some discrepancies in student groups.

When reviewing the CASE English/language arts assessment data we see a score gap between the baseline assessment for the control and experimental group with the experimental group scoring over 5 percentage points higher on the initial assessment. The second assessment that was completed after the introduction of the independent variable showed that both groups did show some growth. However, even though the

control group exceeded the experimental group's growth by 2.366 points, this still did not raise the control group's final score to the initial baseline score level of the experimental group.

This performance discrepancy is also found in reviewing the i-Ready Math Diagnostic scores. Similar to the CASE scores, the experimental group's initial i-Ready scale score average was 17.18 scale score points higher than the control group. In contrast to the CASE data, however, the growth seen by the control group did raise them to a scale score that was comparable to the initial baseline score of the experimental group. Still, the experimental group's growth resulted in these students having an average score over 15 scale score points higher than the control group.

This data indicates that the student make-up in the classrooms of teachers in the experimental group were on average higher performing students from the beginning of the study. Therefore, it could be argued that these students would be expected to show a smaller amount of growth than the initially lower performing control group students. This idea could be further supported considering both the short length of the study and the regression towards mean principle found in statistics. Future research in this area should consider lengthening the time between assessments from one semester to an entire year to provide additional time for the higher performing students to show growth.

The nature of education itself which emphasizes showing a teacher's improvement through student assessment gains may be another factor that impacted the results. A key component of the OKR process is the focus on articulating with specificity what one is trying to achieve and clarity on the steps to reach that achievement level (Doerr, 2018). This closely aligns to a key characteristic of deliberate practice which also

points to improved performance (Ericsson et al., 1993). However, Dunn and Shriener (1999) emphasize the differences between teaching and other domains where deliberate practice has been studied. In this study, teachers created OKRs which may have addressed their behaviors and practices, but the improvement metric is gauged by the student's performance. Future studies in this area may benefit from analyzing the connection between teachers who lead students in setting their own OKRs and the student outcome data.

A final discussion point can be found around the writing and focus of the OKR goals themselves. Doerr (2018) points out that, "an optimal OKR system frees contributors to set at least some of their own objectives and most or all of their key results" (p. 88). This is supported by Locke and Latham (2012) who also connect participation in goal setting versus being assigned goals to individuals setting higher goals and increased performance. In this study, however participants had almost complete control over both the objectives and key results they set instead of some control as mentioned above. Participants did receive feedback on the OKR, but ultimately, they had primary control over the focus area of the goal. Because of this, only 38 percent, or 8 of the 21, of the OKRs tied directly to the student performance outcomes being measured. Assisting teachers in correctly identifying and writing their OKRS is comparable to need identified by Brown et al. (2016) who suggests teachers also need assistance in writing PLC SMART Goals. As he states, "writing SMART goals appears to be a complex problem-solving process for many teachers" (p. 629). Based on this, it is understandable that writing effective OKRs can be equally as complex for teachers especially around identifying the correct objective to be targeted for improvement. Future research in this

area could examine the results when the OKR process more closely follows the optimal system described by Doerr above where participants have control over the key results but have more defined constraints over the objectives they are trying to achieve.

### *Teacher Motivation*

Regardless of the impact on student growth, the literature clearly pointed to the positive impact goal setting and the OKR process should have on teacher motivation. Camp (2017) showed that teachers specifically found goal setting as a positive investment of time and that it led to increased perceptions that their teaching had improved. Furthermore, Camp (2017) indicated that even teachers who did not make significant progress on their goal expressed that the process was valuable. This alignment between increased motivation and goal setting is clearly supported by the research of Ryan and Deci (2000), Bandura (2000), Pink (2009), and Locke and Latham (2019). In addition, Doerr (2018) specifically ties OKRs to motivational improvements in the companies implementing the process.

Based on this abundance of literature, the hypothesis expected our experimental group to show a statistically significant difference in their motivational levels. However, as previously discussed this difference did not exist. In addition, in each of the eight questions, the control group had a higher average mean score than the experimental group. In this analysis though we see a population group that is very closely aligned in their responses around motivation. On the 1 to 4 scale, the lowest average mean score of the entire population was 3.45. When separated into the control and experimental group, the lowest average score was 3.38 for the experimental group and 3.5 for the control group. This data indicates that the group as a whole is highly motivated.

Further analysis confirms the strong motivation displayed by the entire population. Of the 40 participants who completed the survey only two scored any of the eight questions as a 2, disagree. These two participants represented one member of the control group and one member of the experimental group. In total, the survey provided 320 data points, and of the 320 only seven data points represented anything below a 3, agree, or 4, strongly agree. Notably, no survey participant rated any item as a 1, strongly disagree. The seven data points that showed a disagree response represented only 2.18% of the total responses given. Meaning over 97% of responses were in the agree or strongly agree category. Due to the high motivational scores of the group as a whole, it is not surprising that a statistically significant difference was not found.

#### *Environment*

Doerr (2018) clearly states that it takes more than just creating OKRs to ensure that the work leads to the desired improvements. He emphasizes that “goals cannot be attained in a vacuum” (p. 212) and points to ensuring there is a correct culture in place to see this succeed. Risk-taking, transparency, and a willingness to be open to feedback are all critical elements in the OKR process. These elements cannot occur if there is not a sense of trust between the individual creating the OKRs and the supervisor or coach who is working with the employee. Therefore, schools should be cautious about implementing the OKR process and expecting results if there is not a culture of openness and feedback already in place. Ultimately, teachers may be inclined to not take the risk required or be open to the transparency if they do not feel the support will be in place from those they are working with.

When a trusting and supporting environments are in place, psychological safety becomes more prevalent in the organization. Edmondson (2004) emphasizes that when people feel psychologically safe, they are more likely to engage in behaviors such as seeking feedback, seeking help when needed, innovating and expanding boundaries.

In this study, we can look at the contrast between the control and experimental groups in terms of years in service at the school when considering the impact of environment. Half of the experimental group were in their first year at the school which was double the number of those in the control group. Considering the study started at the beginning of the year, it would be expected that the experimental group did not have the time to feel as psychologically secure or have a deep feeling of safety that those in the control group may have felt because they were much newer to the school environment.

### *Reflection*

A thorough review of the teacher performance data, student outcome data, and teacher motivation data paints a picture of the two groups that may not have been on initial equal footing and therefore the results should be viewed in light of this discrepancy. Although both the control and experimental groups showed participants who were highly motivated, there were some notable key differences. Quantitatively, the students in the experimental group were not only higher achieving initially but were also taught by teachers with initially higher evaluation scores. The context of the year in which the study was conducted most likely played a key role specifically in the growth seen on the student outcome data for the lower performing students.

This study was conducted during the fall semester of the 2020-2021 school year in which all educational institutions were grappling with how to respond to the national



Coronavirus 2019 (COVID 19) pandemic. The response plan implemented by the Central School District where Southwest Elementary is located included aspects that greatly impacted the way in which students not only received instruction but also impacted how the assessment data was collected.

The school year started with students operating in a hybrid instructional model for the first half of the semester. This meant kindergarten through 2<sup>nd</sup> grade students were being educated at school in-person each day. However, all kindergarten through 2<sup>nd</sup> grade parents had the option of keeping their children at home and allowing them to attend school virtually. 3<sup>rd</sup> through 6<sup>th</sup> graders started the year attending school virtually and learning in the virtual environment. However, 3<sup>rd</sup> – 6<sup>th</sup> grade parents could apply for waivers allowing their children to attend in-person if they lacked internet access or appropriate childcare. The result of this was that there was a mix of students learning at home and in school. Additionally, the option allowing parents to waiver out of the hybrid model meant that the majority of classrooms had students attending both in-person and online. Not only were teachers having to learn how to effectively deliver instruction in an online environment, but many had to do this while also working with students who were in-person in their classrooms.

The implications of the beginning of the year in a hybrid model meant that beginning of the year assessments were done in multiple environments. In-person students took the CASE and i-Ready Diagnostic in the traditional school setting where strong testing protocols and standard administration helped ensure valid and reliable data points. In contrast, virtual students took the assessment online while not under the direct supervision of a teacher in the room with them. Although the intent and desire was to

monitor student activity while taking the assessment, it cannot be guaranteed that the students took the assessments without the assistance of others in the home. In addition, the home environment could not be standardized and guaranteed to be free from distractions like the classroom environment.

By the second half of the semester, the school had moved away from the hybrid model to a fully in-person model. However, parents still had the opportunity to apply for waivers to keep their children virtual. The result of this was that there were less virtual students and less hybrid classes, but the school continued to have classrooms that were operating on a fully virtual model. When the second set of student assessments were given at the end of the semester, they were again delivered in the same method as the first assessments. In-person students participated in the assessments in traditional classroom settings that were designed to be standardized, consistent and free from distraction. Virtual students took the assessments in the home environment where the teacher had less control of the testing environment. In addition, the movement of students from virtual to in-person meant that some students may have taken their first test virtually and the second one in-person. When considering an individual student's growth or lack of growth seen, this change in method must be considered as a potential impact on validity.

The environment of the school during the research study period also impacted the teacher evaluation system in which the teacher performance data was collected.

Although the evaluators were using the same TEAM rubric as in previous years, the tool was being used in a variety of new settings. Some teachers were observed teaching completely virtually over platforms such as ZOOM while others were observed in traditional in-person settings. Even for those being observed in-person, their classroom

setting and the effective strategies they may have used in the past were forced to be adjusted due to safety protocols such as social distancing.

Finally, a review of the previous year may shed light on the growth seen by the control group who had an initial lower average score in the student outcome measures. The 2019-2020 school year ended with a school shut down in March. This was a full two and a half months before the scheduled end of school. Although, the school did provide at-home learning activities, there was not in-person instruction for the last part of the year. Due to this, the school put a focus on working to catch-up the students most behind when the 2020-2021 school year commenced. This means that they targeted additional instruction and intervention supports to those who were most impacted by the early school closure the previous year. In addition, this meant that a larger amount of time than normal was focused on review when the 2020-2021 school year began. Both of these factors would benefit the initially lower achieving students found in the control group over the initially higher achieving students found in the experimental group.

In reflecting on the data found in this study, it must be examined in light of the unprecedented way schools were operating during the study. Overwhelmingly, the research and literature pointed to a hypothesis that would strongly support the connection between improved teacher and student outcomes with the implementation of the OKR process. The lack of evidence found in this study should not be cause to disregard the practice, but instead, it should be reexamined in times when data collection is more standardized, and the environment is more stable.

## **Limitations**

This study included several limitations which could have impacted the results of the research. Addressing these limitations would provide future researchers the opportunity to reexamine the findings and implications of the study.

As previously addressed, this study was conducted during the COVID-19 pandemic and all results must be viewed through this lens. Although instruction, assessments, and teacher evaluation were all occurring during this time, it was often occurring in new and non-standardized ways. A major impact of the hybrid model of instruction was the limitation of an inability to fully control the environments of the study. Although real world research always has limitations in the controlling of environments, the specifics of the district's response to the COVID-19 pandemic magnified this limitation greatly.

The sample size and longitude of the study can be seen as limitations to address in future research. The population was isolated to a specific school over a single semester of the school year. Increasing the population size and the number of teachers in the control and experimental groups would provide additional data to assist in answering the research questions.

Increasing the length of time of the study from one semester to an entire school year would allow for additional student growth to occur. Ideally, this study would be conducted over a longer period of time to better encompass the intent of the OKR process. Although Doerr (2018) states that OKRs can be made as often as quarterly, the impact may best be seen when results are viewed over an extended period of time.

Finally, a lack of previous research on implementing OKRs in the school setting served as a limitation in that there were not previous studies to examine before beginning this study. As noted in the discussion above, this study has provided additional guidance to future researchers on how to better implement the OKR process to impact the unique nature of the school setting.

### **Implications**

As previously discussed, the literature around goal setting and OKRs specifically led to the hypothesis that implementing this practice would lead to increased teacher and student performance. However, the results of this study did not support this idea. However, there are still areas to consider in light of this information and the study does point to important implications for practice, policy, and future research in this area.

#### *Implications for Practice*

The research and literature previously discussed in this study points to the value and importance of including goal setting protocols in improvement models, and despite the results of this study, this should not exclude education. There are several implications for practice that should be closely considered in light of this research.

First, Locke and Latham (2019) discuss how goals provide clear direction for the behaviors as you work to attain your objectives. Having a clear direction allows for specificity in knowing what one is trying to achieve. The OKR process builds on this by not only identifying what you are trying to achieve, but it also articulates the specific steps you will take to help reach this goal. Teaching is an incredibly complex profession, and educators are consistently pulled in multiple directions all of which are competing for their time and attention. By clearly participating in a goal-setting process such as OKRs,

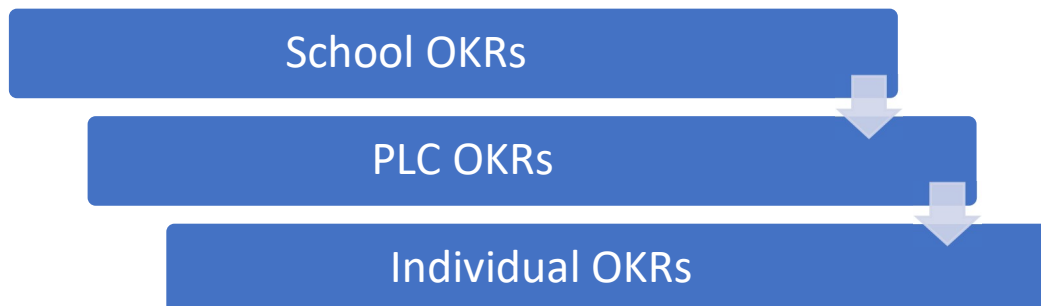
educators are assisted in narrowing down what is urgent and what needs to be done from other less important or impactful tasks that may call for their attention.

Locke and Latham (2019) additionally discuss how goals not only provide direction for the behavior, but that they also serve as a motivator for performance. This increased motivation results in both our persistence in reaching our goals and the intensity with which we work. Increasing the internal motivation of teachers is becoming increasingly important as outside forces such as low teacher pay and increased workloads is often resulting in decreased teacher motivation and moral. Ryan and Deci (2020) emphasize the benefits of strong intrinsic motivation as not only increased performance but enhanced engagement in the learning process. Therefore, teachers should strongly consider implementing some type of goal setting process as part of their work.

A second implication is the potentially positive impact that OKRs could have on student outcomes when they are implemented directly with students. Locke and Latham (2020) indicate that goals can impact and affect a participant's persistence in meeting a task. As students continue to face increased performance expectations through their schooling, it will be critical that they also build their ability to persist through challenges. As previously discussed, the goal setting protocol also contributes to increasing the intrinsic motivation by helping ensure their universal needs of competency, autonomy, and relatedness are met (Deci & Ryan, 2008a). As Pink (2009) articulates, the results of this increased student motivation can lead to "greater conceptual understanding, better grades, enhanced persistence at school in sporting events, higher productivity, less burnout, and greater levels of psychological well-being" (p. 89). All of these

characteristics are critical as we seek as educators to empower the lives of the students we serve.

Finally, there is a strong implication for administrators and district leaders who oversee the improvement process of teachers. It is critical that teachers have a voice in their personal improvement process regardless of the role goal setting plays in it. The OKR process is built on the idea of collaboration around goals, on-going feedback, and offering intentional support to the person working to meet the goals. Doerr (2018) points out that OKRs support this basic truth of human nature, “When people choose a course of action, they are more likely to see it through” (p. 25). In a school environment, implementing OKRs may prove more valuable when viewed within an established PLC process. This could include PLC teams first creating team OKR statements followed by individuals creating their own OKRs that support the team’s work. By aligning OKRs to the work and goals of the PLC, two key attributes of the OKR process are emphasized. First, this will assist teachers in ensuring their OKRs are relevant to the goals of the organization and that they are focused on the right work. Secondly, this will build off the public sharing and transparency component of the OKR process which Doerr (2018) underscores as critical to success and serves as a way for teams to “attack trouble spots wherever they surface” (p. 89). Furthermore, the increased transparency leads to increased goal commitment (Camp, 2017). The figure below illustrates how the OKR process can filter from the school level to the individual level to build relevance and coherence across the school. The role of the teacher is key, and it is incumbent on the leader to ensure the teacher is an active participant in the improvement process and the current school structures help set the teacher up for success.

*Figure 5**OKR Filtering Process of Alignment**Implications for Policy*

This research study does not support the implementation of policy around the OKR system. However, the literature does support the implementation of key components of the OKR process. Specifically, these elements should be considered in current policies around teacher evaluation. Research has indicated that the primary purposes for teacher evaluation are both accountability or quality assurance and teacher support and development (Danielson 2001, 2010). However, (Danielson, 2001) argues that previous evaluation systems have failed at accomplishing either goal. If we are going to ensure we have a system where teacher development is equally as important as accountability, then we must make certain that teachers are actively engaged in the evaluation process. Future policies around teacher evaluations should guarantee that educators have the ability to provide some level of control in identifying the areas which they want to target for improvement. As Danielson (2010) states, “learning is done by the learner through the process of active intellectual engagement” (p. 38).



The failure of current evaluation systems to meet either of their stated goals may indicate the need of using separate tools to address each purpose. Specifically, it should be considered that evaluation systems are designed and used for the sole purpose of accountability while allowing alternative models such as the OKR process to be used for improvement models. By divorcing the improvement structure from the accountability structure may help ensure teachers embrace improvement and the willingness to fail that often accompanies the work of improvement.

In addition to intrinsic motivation that occurs when teachers have a level of autonomy in their improvement, they are also more prepared to participate in deliberate practice aimed at improving the area of need. Ericsson et al. (1993) differentiates between participation and deliberate practice and clearly links deliberate practice to increased performance. Therefore, it is clear that although OKRs may not be fully supported as a policy recommendation, we should be reviewing the role educators plays in their own improvement models if we are truly sincere about using teacher evaluations as a method to improve instruction.

#### *Implications for Future Research*

The limitations presented in this study, specifically around the COVID-19 pandemic, present the opportunity for future research to gain a greater understanding of the impact of implementing the OKR model in the school setting. Future researchers should be encouraged to attempt this research again when the environmental factors are not as uncontrollable as they were in this study. Specifically, it would be beneficial to attempt this in a setting where all students are in similar settings with similar controls over data collection and testing environments. Controlling for these areas could produce

results that are more aligned with the literature and theories used to support the initial work.

In addition, future researchers should examine the use of OKRs specifically with students to better understand the impact of this process in the educational environment. As stated previously, education is unique in that the teacher's performance is evaluated not just on their own performance but also the performance of the students who learn from them. Therefore, the connection could be made that a more significant impact is possible by including students in the OKR process since they are the ones who are directly involved in producing the outcome measurements.

### **Conclusion**

Dylan Wiliam (2018) reminds us that the "job of teaching is so difficult, so complex, that one lifetime is not enough to master it" (p. 29). Despite this challenge, it is incumbent upon all educators to work each day to systematically improve in their craft and understanding of how to improve student outcomes. Unlike other professions where undesired results may have a financial downside or end with outcomes that are personally detrimental, undesired results in education have very different long-term impacts. In education, when an employee does not improve, they are not the ones to pay the price. Instead, the impact is felt by classrooms of students with whom they have been entrusted to educate, grow, and empower. This great responsibility makes it not only desirable to improve but a moral imperative for all educators.

This work sought to take a proven goal setting improvement model used in business and evaluate the impact it would have in the educational setting. It was grounded in research-based theories of goal setting and human motivation as well as

supported by literature on effective teaching evaluations, deliberate practice, and environments where this work could best thrive.

Ultimately, the findings of this study did not support the hypothesis that implementing Doerr's (2018) Objectives and Key Results process would lead to increased teacher performance, student learning outcomes and teacher motivation. However, unprecedented limitations and environmental factors were largely at play that impacted the research process.

Regardless of these specific results, it is still strongly recommended that goal-setting processes play a part in the improvement model for teachers and schools. In addition, it is critical that teachers have a strong voice in identifying the areas in which they are seeking to improve. By doing this we are not only providing the opportunity to increase the motivation of teachers, but we are also providing support that will assist them in engaging in the deliberate practice that will help build their expertise.

As we continue to navigate through the COVID-19 experience we find ourselves in as a country, we must be aware that our students are perhaps the ones who have been most impacted. Their learning has been interrupted, their engagement in school activities has been decreased or stopped altogether, and the gap for our most school-dependent students has certainly grown. All of this provides even more urgency around the work educators do each day. What we have done in the past will most likely not be enough for what is needed in the future. Despite the fact that this research study did not show hypothesized positive results, improvement is still the imperative, and we must look for new and proven ways to do this. It is essential that we understand that "if we continue to

seek improvement in the ways we have always done, we are likely to continue to get what we have always gotten” (Bryk et al., 2017, p. 6).

## REFERENCES

- Adaptive Assessment That Dives Deep. (n.d.). Retrieved June 10, 2020, from <https://www.curriculumassociates.com/products/i-ready/i-ready-assessment/diagnostic>
- Baard, P. P., Deci, E. L., & Ryan, R. M. (2004). Intrinsic need satisfaction: a motivational basis of performance and well-being in two work settings. *Journal of Applied Social Psychology, 34*(10), 2045-2068.
- Bandura, A. (1997). Self-efficacy. *Harvard Mental Health Letter, 13*(9), 4-6.
- Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science, 9*(3), 75-78. doi: 10.1111/1467-8721.00064
- Bandura, A. & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*(1), 87-99. doi: 10.1037/0021-9010.88.1.87
- Bell, J. (1999). Peter F. Drucker (b. 1909): management's modernizer. *The Journal of Business Strategy, 20*(5), 35-36.
- Brown, G., Leonard, C., and Arthur-Kelly, M. (2016). Writing SMARTER goals for professional learning and improving classroom practices. *Reflective Practice, 17*(2), 621-635. doi: 10.1080/14623943.2016.1187120
- Bryk, A. S., Gomez, L. M., Grunow, A. & LeMahieu, P. G. (2017). *Learning to Improve: how America's schools can get better at getting better*. Cambridge: Harvard Education Press.
- Camp, H. (2017). Goal setting as teacher development practice. *International Journal of Teaching and Learning in Higher Education, 29*(1), 61-72.

- Chamberlin, J. (2011). Who put the ‘ART’ in SMART goals? *Management Services*, 55(3), 22-27.
- Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic review: the relationship between clinical experience and quality health care. *Annals of Internal Medicine*, 142(2), 260-273.
- Cresswell, J. W. & Cresswell J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5<sup>th</sup> ed.). Thousand Oaks: Sage.
- Connally, K. & Tooley, M. (2016). Beyond ratings re-envisioning state teacher evaluation systems as tools for professional growth. *New America*. Retrieved from Tennessee Department of Education TEAM website: <https://team-tn.org/research/>
- Cosner, S. (2009). Building organizational capacity through trust. *Educational Administration Quarterly*, 45(2), 248-291. doi: 10.1177/0013161x08330502
- Coughlan, E. K., Williams, A. M., McRobert, A. P., & Ford, P. R. (2013). How experts practice: a novel test of deliberate practice theory. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 40(2), 449-458. doi: 10.1037/a0034302
- Curriculum Associates. (n.d.). *i-Ready Diagnostic: Linking Study with TNReady*. Retrieved from: <https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-diagnostic-assessments-linking-study-overview-tennessee-2020.pdf>
- Danielson, C. (2001). New trends in teacher evaluation. *Association for Supervision and Curriculum Development*, 58(5), 12-15.
- Danielson, C. (2010). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35- 39.

- Darling – Hammond, L. (2014). One piece of the whole: teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator*, 38(1), 4-13.
- Darling-Hammond, L. & Goldberg, M.F. (2001). An interview with Linda Darling-Hammond: balanced optimism. *The Phi Delta Kappan*, 82(9), 687-690.
- Davis, J., Lampley, J. H., & Foley, V. (2016). The relationship between growth scores and the overall observation ratings for teachers in a public school system in Tennessee. *Journal of Learning in Higher Education*, 12(1), 45 – 51.
- Deci, E. L. & Ryan, R. M. (2008a). Facilitating optimal motivation and psychological well being across life's domains. *Canadian Psychology*, 49(1), 14-23.  
doi: 10.1037/0708-5591.49.1.14
- Deci, E.L. & Ryan, R.M. (2008b). Self-determination theory: a macrotheory of human motivation, development, and health. *Canadian Psychology*, 49(3), 182-185.  
doi: 10.1037/a0012801
- Dirks, K. T. & Ferrin, D. L. (2002). Trust in leadership: meta-analytic findings and implications for research and practice. *Journal of Applied Psychology*, 87(4), 611-628.
- Doerr, J. (2018). *Measure what Matters*. New York: Portfolio/Penguin.
- DuFour, R. [Richard], DuFour R. [Rebecca]. (2008, September 10). Team SMART goals vs. smart students [Blog Post]. Retrieved from  
<https://www.allthingsplc.info/blog/view/24/Team+SMART+Goals+vs.+Smart+Students>

- DuFour, R. [Richard], DuFour R. [Rebecca], Eaker, R., Many, T.W., Mattos, M. (2016). *Learning by doing A handbook for professional learning communities at work* (3<sup>rd</sup> ed.). Bloomington: Solution Tree Press.
- Dunn, T. G., & Shriener, C. (1999). Deliberate practice in teaching: what teachers do for self-improvement. *Teaching and Teacher Education, 15*(6), 631-651.  
doi: 10.1016/S0742-051c(98)00068-7
- Dall'Alba, G. (2018a). Reframing expertise and its development: a lifeworld perspective. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds). *The Cambridge handbook of expertise and expert performance, 2<sup>nd</sup> ed.* (33-39). Cambridge University Press. doi: 10.1017/9781316480748.003
- Eaker, R., Keating, J. (2012). *Every school, every team, every classroom District leadership for professional learning communities at work*. Bloomington: Solution Tree Press.
- Edmondson, A. (2004). Psychological safety, trust, and learning in organizations: a group-level lens. In R. Kramer & K. Cook (Eds.), *Trust and distrust in organizations: dilemmas and approaches* (pp. 239-272). New York: Russel Sage.
- Edmondson, A. (2019). The role of psychological safety: maximizing employee input and commitment. *Leader to Leader, 92*, 13-19. doi: 10.1002/ltl.20419
- Edmondson, A., Higgins, M., Singer, S., & Weiner, J. (2016). Understanding psychological safety in health care and education organizations: a comparative perspective. *Research in Human Development, 13*(1), 65-83.  
doi: 10.1080/15427609.2016.1141280



Elementary and Secondary Education Act of 1965, H.R. 2362, 89<sup>th</sup> Cong. (1965)

(enacted)

Ericsson, K. A. (2018a). An introduction to the second edition of the Cambridge handbook of expertise and expert performance: its development, organization, and content. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds). *The Cambridge handbook of expertise and expert performance, 2<sup>nd</sup> ed.* (3-20). Cambridge University Press. doi: 10.1017/9781316480748.001

Ericsson, K. A. (2018b). The differential influence of experience, practice, and deliberate practice on the development of superior individual performance of experts. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds). *The Cambridge handbook of expertise and expert performance, 2<sup>nd</sup> ed.* (745-769). Cambridge University Press. doi: 10.1017/9781316480748.038

Ericsson, K. A. & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725-747.

Ericsson, K. A. & Harwell, K. W. (2019). Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: why the original definition matters and recommendations for future research. *Frontiers in Psychology, 10*, 1-19. doi: 10.3389/fpsyg.2019.02396

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

- Ericsson, K. A., & Lehmann, A. C., (1996). Expert and exceptional performance: evidence and maximal adaptations to task constraints. *Annual Review of Psychology*, 47(1), 273-305. doi: 10.1146/annurev.psych.47.1.273
- Ford, R.C., Mclaughlin, F.S., and Nixdorf, J. (1980). Ten questions about MBO. *California Management Review*, 23(2), 88 – 94. doi: 10.2307/41164922
- Garvin, D. A., Edmondson, A. C., & Gino, F. (2019). Is yours a learning organization? [Special Issue]. *Harvard Business Review*, 86-93.
- Geppert, L. (2000). Andy Grove [Profile]. *IEEE Spectrum*, 37(6), 35-38.  
doi:10.1109/MSPEC.2000.846095
- Gladwell, M. (2008). *Outliers: the story of success*. New York: Little, Brown and Co.
- Gopalan, M., Rosinger, K., & Ahan, J.B. (2020). Use of quasi-experimental research designs in education research: growth, promise, and challenges. *Review of Research in Education*, 44(1), 218-243. doi: 10.3102/0091732X20903302
- Greenwood, R.G. (1981). Management by objectives: as developed by Peter Drucker, assisted by Harold Smiddy. *Academy of Management Review*, 6, 225-230.
- Handford, V. & Leithwood, K. (2013). Why teachers trust school leaders. *Journal of Educational Administration*, 51(2), 194-212. doi: 10.1108/09578231311304706
- Hattie, J. (2012). *Visible learning for teachers: maximizing impact on learning*. New York: Routledge.
- Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the widget effect: teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(5), 234 – 249. doi: 10.3102/0013189X17718797

- Lapidot, Y., Kark, R., & Shamir, B. (2007). The impact of situational vulnerability on the development and erosion of followers' trust in their leader. *The Leadership Quarterly, 18*(1), 16-34. doi: 10.1016/j.leaqua.2006.11.104
- Lencioni, P. (2002). *The five dysfunctions of a team: A leadership fable*. San Francisco: Jossey-Bass.
- Locke, E.A. (1968). Toward a theory of task motivation and incentives. *Organizational Power & Human Performance, 3*(2), 157-189.  
doi: 10.1016/0030-5073(68)90004-4
- Locke, E. A. (2000). Motivation, cognition, and action: an analysis of studies of task goals and knowledge. *Applied Psychology: An International Review, 49*(3), 408 – 429. doi: 10.1111/1464-0597.00023
- Locke, E. A., & Latham, G. P. (1990). Work motivation and satisfaction: light at the end of the tunnel. *Psychological Science, 1*(4), 240-246.  
doi: 10.1111/j.1467-9280.1990.tb00207.x
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *The American Psychologist, 57*(9), 705-717.
- Locke, A. L. & Latham, G. P. (2019). The development of goal setting theory: a half century retrospective. *Motivation Science, 5*(2), 93-105.  
doi: 10.1037/mot0000127
- MacLeod, L. (2012). Making SMART goals smarter. *Physician Executive, 38*(2), 68 – 72.

- McBride, M. & Skau, K. G. (1995). Trust, empowerment, and reflection: Essentials of supervision. *Journal of Curriculum & Supervision*, 10(3), 262-277.
- McGreal, T. L., (1980). Helping teachers set goals. *Educational Leadership*, 37(4), 414-419.
- Moye, M. J., Henkin, A. B., & Egley, R. J. (2004). Teacher-principal relationships: exploring linkages between teacher empowerment and interpersonal trust. *Journal of Educational Administration*, 43(3), 260-277. doi: 10.1108/09578230510594796
- Newman, A., Donahue, R., & Eva, N. (2017). Psychological safety: a systematic review of literature. *Human Resource Management Review*, 27(3), 521-535.  
doi: 10.1016/j.hrmr.2017.01.001
- No Child Left Behind Act of 2001, H.R. 1, 107<sup>th</sup> Cong. (2001) (enacted).
- O'Neill, J. (2000). SMART goals, SMART schools. *Educational Leadership*, 57(5), 46-50.
- Page, L., Boysen, S., & Arya, T. (2019). Creating a culture that thrives: fostering respect, trust, and psychological safety in the workplace. *OD Practitioner*, 51(1), 28-35.
- Pink, D. H. (2009). *Drive: the surprising truth about what motivates us*. New York: Riverside Books.
- Ryan, R. M. & Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54-67.  
doi: 10.1006/ceps.1999.1020

- Ryan, R. M. & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. *Contemporary Educational Psychology*.  
doi:10.1016/j.cedpsych.2020.101860
- Ryan, R. M. & Powelson, C. L. (1991). Autonomy and relatedness as fundamental to motivation and education. *Journal of Experimental Education*, 60, 49-66.
- Tedlow, R.S. (2005). The education of Andy Grove. *Fortune*, 152(12), 116-138.
- TE21, Inc. CASE Assessments Research Overview. (n.d.). Retrieved June 10, 2020, from [https://www.te21.com/pages/page.asp?page\\_id=359422](https://www.te21.com/pages/page.asp?page_id=359422)
- TE21, Inc. CASE Benchmark Assessments. (n.d.). Retrieved June 10, 2020, from <https://www.te21.com/case>
- Tennessee Department of Education. (2011). *Tennessee Educator Acceleration Model (TEAM) Performance Standards Research Base* Retrieved from Tennessee Department of Education TEAM website: <https://team-tn.org/wp-content/uploads/2013/08/TEAM-Rubric-Research-Base.pdf>
- Tennessee Department of Education. (2016). *Teacher and administrator evaluation in Tennessee a report on year 4 implementation*. Retrieved from Tennessee Department of Education TEAM website: <https://team-tn.org/research/>
- Tennessee Department of Education. (2018). *Tennessee Educator Acceleration Model TEAM Teacher Evaluation Evaluator Handbook*. Retrieved from Tennessee Department of Education TEAM website: [https://team-tn.org/wp-content/uploads/2013/08/2018-19-TEAM-Teacher-Evaluator-Handbook\\_1002181.pdf](https://team-tn.org/wp-content/uploads/2013/08/2018-19-TEAM-Teacher-Evaluator-Handbook_1002181.pdf)

- Tennessee Educator Research Alliance. (n.d.). Retrieved February 13, 2021, from [https://peabody.vanderbilt.edu/TERA/Tennessee\\_Educator\\_Survey.php](https://peabody.vanderbilt.edu/TERA/Tennessee_Educator_Survey.php)
- Tennessee Educator Survey. (n.d.). Retrieved February 13, 2021, from <https://www.tn.gov/education/data/educator-survey.html>
- Tschannen-Moran, M. & Hoy, W. K. (2000). A multidisciplinary analysis of the nature, meaning, and measurement of trust. *Review of Educational Research*, 70(4), 547-593. doi: 10.3102/00346543070004547.
- Tucker, M. S. & National Center on Education and the Economy (2014). Fixing our national accountability system. *National Center on Education and the Economy*. Retrieved from National Center on Education and the Economy website: <http://ncee.org/accountability/>
- U.S. Department of Education. (2009). *Race to the top executive summary*. Retrieved from U.S. Department of Education website: <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- Vohra, N., and Mukul, K. (2009). Relevance of Peter Drucker's work: Celebrating Drucker's 100<sup>th</sup> birthday. *Vikalpa: The Journal for Decision Makers*, 34(4), 1 – 7. doi: 10.1177/0256090920090401
- Ward, P., Hodges, N. J., Starkes, J. L., & Williams, A. M. (2007). The road to excellence: deliberate practice and the development of expertise. *High Ability Studies*, 18(2), 119-153. doi: 10.1080/13598130701709715
- Weisberg, D., Sexton, S., Mulhern, J., Kelling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: our national failure to acknowledge and act on differences in teacher effectiveness*. Retrieved from The New Teacher Project

website: <https://tntp.org/publications/view/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness>

Whitener, E. M., Brodt, S. E., Korsgaard, M. A., & Werner, J. M. (1998). Managers as initiators of trust: an exchange relationship framework for understanding managerial trustworthy behavior. *The Academy of Management Review*, 23(3), 513-530.

Wiliam, D. (2018). *Embedded formative assessment* (Second ed.). Bloomington: Solution Tree Press.

Willingham, D. T. (2009). *Why don't students like school?: A cognitive scientist answers questions about how the mind works and what it means for the classroom*. San Francisco: Jossey-Bass.

Zahra, S.A. (2003). An interview with Peter Drucker. *Academy of Management Executive*, 17(3), 9 – 12. doi: 10.5465/AME.2003.10954665