

Florida Institute of Technology

Scholarship Repository @ Florida Tech

Theses and Dissertations

3-2018

Using the Operant Model of Effective Supervision to Predict Employee Engagement for Leaders in a Hospital System

David Patrick Kelley III

Follow this and additional works at: <https://repository.fit.edu/etd>



Part of the [Applied Behavior Analysis Commons](#)

Using the Operant Model of Effective Supervision to Predict Employee
Engagement for Leaders in a Hospital System

by

David Patrick Kelley III

A dissertation submitted to the School of Behavior Analysis of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Behavior Analysis

Melbourne, Florida
March, 2018

The undersigned committee,
having examined the attached dissertation,
“Using the Operant Model of Effective Supervision to Predict Employee
Engagement for Leaders in a Hospital System,” by
David Patrick Kelley III
hereby indicates its unanimous approval.

Nicole Gravina, Ph.D.
Assistant Professor
School of Behavior Analysis

David Wilder, Ph.D.
Professor
School of Behavior Analysis

Nicholas Weatherly, Ph.D.
Associate Professor
School of Behavior Analysis

Xinxuan Che, Ph.D.
Assistant Professor
School of Psychology

Mary Beth Kenkel, Ph.D.
Professor, Dean
College of Psychology and Liberal Arts

Abstract

Title: Using the Operant Model of Effective Supervision to Predict Employee Engagement for Leaders in a Hospital System

Author: David Patrick Kelley III

Advisor: Nicole E. Gravina, Ph.D.

Leadership has been a popular topic of conversation for decades. Despite an abundant amount of attention the topic has received, the literature seems to fall short when describing what effective leaders do in a day to day context. A theory introduced by Komaki (1986), called the Operant Model of Effective Supervision (OMES), was designed to answer the tough questions about leadership that have been left unanswered. Research on the theory has resulted in multiple methods to accurately measure leadership behavior within an operant paradigm. One of these methods is an in-basket assessment that has been shown to reliably capture day to day leadership behavior in about one hour. The present study sought to assess whether this tool was predictive of leadership effectiveness by comparing scores on the assessment

of 47 individuals in a leadership role with an employee engagement score, which captures follower attitudes and is reliably linked to organizational performance. The study found that leaders who spend more time providing antecedents to employees have significantly lower employee engagement scores than leaders who spend less time providing antecedents.

Table of Contents

Introduction	1
Leadership Criteria	2
Operant Conditioning Theory	3
The Operant Model of Effective Supervision	8
Research on the Operant Model	14
Operant Supervisory In-Basket Assessment (OSIBA).....	24
Employee Engagement.....	31
Creating Employee Engagement	37
Method	40
Participants and Setting	40
Design.....	42
Operant Supervisor In-basket Assessment (OSIBA)	43
OSIBA Categories and Scoring.....	43
Scoring	44
Interrater Agreement	45
Criterion: Employee Engagement	46
Procedure.....	47
Analysis.....	48
Results	48
Interrater Percentage Agreement.....	48
Descriptive Statistics for In-basket Categories and Subcategories	50
In-basket Category and Subcategory Relationships	52
Demographic Variables Relation to In-basket Category Scores	54
Descriptive Statistics of the Criterion Variable.....	56
Demographic Variables Relation to Employee Engagement	56
Hypothesis Testing	58
Discussion	61
Limitations	70
Future Research.....	73
References	78

List of Keywords

Leadership

Operant

Model

Effective

Supervision

Behavior

List of Tables

Table 1. <i>Categories and Subcategories of Behavior in the OSIBA: Definitions and Examples</i>	88
Table 2. <i>Summary of Demographic Information</i>	93
Table 3. <i>Points Possible for Categories and Subcategories of Behavior</i>	94
Table 4. <i>Average Interrater Agreement Scores for Major Categories</i>	95
Table 5. <i>Summary of In-basket Scores</i>	96
Table 6. <i>Intercorrelations Between Major Categories and Subcategories of Behavior in the OSIBA</i>	98
Table 7. <i>Correlations Between Selected Demographic Variables and In-basket Scores</i>	99
Table 8. <i>Descriptive Statistics of Criterion Variable</i>	100
Table 9. <i>Correlations Between Selected Demographic Variables and Employee Engagement Scores</i>	101
Table 10. <i>Correlations Between In-basket Category and Subcategory Scores and Employee Engagement Scores</i>	102
Table 11. <i>Hierarchical Multiple Regression Results Using Antecedents and N Size as Predictors of Employee Engagement Scores</i>	103
Table 12. <i>Hierarchical Multiple Regression Results Using N Size, Monitors, Consequences, and an Interactive Term as Predictors of Employee Engagement Scores</i>	104
Table 13. <i>Hierarchical Multiple Regression Results Using N Size, Monitors, Consequences, and Antecedents as Predictors of Employee Engagement Scores</i>	105
Table 14. <i>Tacking on Traditional Antecedents in Response to an OSIBA Item from a Leaders with High and Low Employee Engagement Scores</i>	106

List of Figures

Figure 1. Components of the operant model of effective supervision. Reprinted from <i>Leadership from an Operant Perspective</i> (p. 26), by J. L. Komaki, 1998, London: Routledge. Copyright 1998 by Judith L. Komaki. Reprinted with permission.	107
---	-----

Introduction

For decades, organizations have struggled with a shortage of qualified leaders (Tucker & Lam, 2014). Despite this continued shortage and a plethora of research, organizations are still working to identify programs for selecting, training, and producing effective leaders (Krapfl & Kruja, 2015). Most researchers and practitioners today acknowledge the importance of leadership for the effectiveness of an organization (Yukl, 2013). To date, it is generally agreed that leadership refers to the process of influencing others (Vroom & Jago, 2007). Daniels and Daniels (2007) propose that a leader's behavior should, "establish the conditions under which all performers will choose to execute the mission, vision, and values of the organization" (p. 4). Houmanfar, Alavosius, Morford, Herbst, and Reimer (2015) add that, "leadership behaviors include effectively communicating the mission and vision of an organization, cultivating a motivated workforce, and ensuring adequate resources for production" (p. 17).

Measuring leadership behavior in the workplace has proven difficult, leaving the behavior of leaders who are effective at "influencing others" a mystery. Similarly, finding universal criteria to serve as a measure of leadership effectiveness has also proven to be a challenge. One assessment tool, the Operant Supervisory In-basket Assessment (OSIBA), has been developed that reliably measures leadership behavior in an efficient manner (Komaki, 1998). However, we

still do not know what specific leadership behaviors drive team performance and follower attitudes. Employee engagement scores have been found to be a reliable universal measure of leadership effectiveness and are highly correlated with business results (Harter, Schmidt, Agrawal, Plowman, & Blue, 2016). Determining which leadership behaviors, as measured by the OSIBA, are predictive of employee engagement scores in an organizational setting is the next logical step for leadership research.

Leadership Criteria

With an abundant amount of research focused on leadership, little is still known as to how effective leaders differ in behavior from ineffective leaders. Follower attitudes and perceptions are often an indicator of leadership effectiveness, which are typically measured with questionnaires or interviews (Yukl, 2013). Another relevant indicator used in research is the extent to which the performance of the team or organization is enhanced and the attainment of goals is facilitated (Bass, 2008; Kaiser, Hogan, & Craig, 2008). Problems exist with each of these criteria. For example, if a leader and her team consistently achieve profit margin goals, but followers are extremely unhappy as a result of her authoritarian style, they may end up leaving or engaging in unsafe or unethical behavior, which will likely result in a hefty cost for the organization over time. As Boris Yeltsin once said, “You can make a throne of bayonets, but you can’t sit on it for very

long” (Murray, 1995). The same holds true for the opposite scenario. If a leader consistently has satisfied associates who wish to stay in their role long term, but his team never achieves a profit, the business is not likely to survive. Capturing both follower attitudes and organization or team performance is preferred (Komaki, 1998), although, this can be a significant challenge.

In addition, acquiring consistent and objective data for performance outcomes across a large sample of leaders is extremely rare and challenging. Even more challenging is measuring leaders based on the same criteria when they are operating in different environments with differing goals. Thus, an easily obtainable criteria which provides an accurate picture of leadership effectiveness in terms of follower attitudes and performance outcomes is needed. Without this measure, it becomes impossible to determine whether certain leadership behaviors are to be classified as helpful versus harmful when exploring leadership effectiveness.

Operant Conditioning Theory

Operant conditioning theory is derived from the work of B. F. Skinner and is also referred to as behavior analysis (Komaki, 1986). The theory focuses on the relationship between the physical environment and the behavior of organisms. In this context, *behavior* can be thought of as, “any portion of an organism’s interaction with its environment that involves movement of some part of the organism” (p. 31) (Johnston & Pennypacker, 2009). Once a specific behavior of

interest has been identified, the environmental stimuli that surround the behavior are of primary interest. Environmental conditions that occur prior to the behavior of interest are known as *antecedents* (Cooper, Heron, & Heward, 2014). For example, if the behavior of interest is employees showing up to work on time, the policies written stating that employees must arrive to work on time would be considered an antecedent. An environmental or stimulus change that follows the behavior of interest is known as a *consequence* (Cooper, Heron, & Heward, 2014). Therefore, how an organization responds to employees showing up on time or not would be considered a consequence (Komaki, 1998). Events such as providing feedback or giving an employee a written warning for showing up late may be considered consequences within this paradigm.

In operant conditioning theory, consequences are the major motivational force, where antecedents play an educational or cuing role (Daniels, 1994; Komaki, 1986; Scott & Podsakoff, 1985; Skinner, 1974). As Skinner (1953) stated, “In operant conditioning, we “strengthen” an operant in the sense of making a response more probable or, in actual fact, more frequent” (p. 65). When a consequence results in an increase in the frequency of a behavior, *reinforcement* has taken place. The consequence responsible for the increase is known as a *reinforcer*. When a consequence results in a decrease in the frequency of a behavior, *punishment* has taken place. The consequence responsible for the decrease is known as a *punisher*

(Cooper, Heron, & Heward, 2014). Under this paradigm, it is important to note that consequences can only affect the occurrence of future behavior. As Cooper, Heron, and Heward (2014) said, “a behavioral consequence affects the relative frequency with which similar responses will be emitted in the future under similar stimulus conditions” (p. 34). This concept has been demonstrated in the research literature numerous times (Schneider, 2012).

In a laboratory study simulating a work environment, researchers found that antecedents such as persuasive influence or announcements alone were insufficient to increase and sustain worker performance (Johnson, 1975). Only when a pay-contingency was implemented did performance improve. In a recent study conducted on sitting posture in the workplace, researchers found that antecedents such as prompts had a minimal impact on safe sitting behavior (Moon & Oah, 2013). When a consequence such as feedback was provided on sitting posture, substantial improvements in safe sitting behavior occurred in all participants.

One review assessed the impact of positive reinforcement in fifty-one well controlled studies in work settings (Komaki, Coombs, & Schepman, 1991). The study found that forty-seven of these studies reported substantial improvement in targeted performance. These findings show that in the literature alone, positive reinforcement appears to have a 92.2 percent success rate for improving performance.

Results from the Gallup Q12 survey, a 12-item survey designed to measure employee engagement has shown that fewer than one in three American workers can strongly agree they have received any praise from a supervisor in the last seven days. Variation in response to this item is responsible for differences in revenue and productivity up to 20% across organizations. To make matters worse, employees who report that they're not sufficiently recognized at their place of work are three times more likely to say they will quit in the next year (Robison, 2006). These results shed light on the important role leadership behavior must play and our need to better understand it.

Despite the existing evidence to suggest that consequence-based strategies are key to improving employee performance, antecedent strategies are still prevalent in the I/O leadership literature. Thorough expectations and inspirational speeches are key elements to transformational leadership theory (Bass, 1985; House, 1977). It is unclear whether or not inspirational speeches impact leadership effectiveness or if transformational leaders also engage in other behaviors such as providing positive feedback consistently. In order to answer these questions, researchers must observe leadership behavior in action to determine which behaviors have the most impact on performance.

One of the most well-known theories of motivation in the workplace which emphasizes the antecedent is goal-setting (Locke & Latham, 1990). That said,

antecedents do play an important role in behavior change, just not a sustainable one on their own. For example, in a workplace where performance goals have never been set, setting a goal might get behavior to improve. The new goal might also make it more likely that an employee comes into contact with new consequences such as feedback in relation to achieving the goal, or the gratification of seeing yourself meet a goal. As Daniels and Bailey (2014) put it, “antecedents might get a behavior started, but only consequences maintain behavior” (p. 131). It is important to note the link between antecedents and consequences as it is often not simply a matter of one or the other. Understanding which specific day-to-day behaviors make a leader most effective in terms of motivating employees has been an ongoing challenge for academics and practitioners.

As a result of this struggle, Komaki, Zlotnick, and Jensen (1986) developed the Operant Model of Effective Supervision to evaluate the specific behaviors that effective leaders engage in. The model, which is grounded in the theory of operant conditioning, focuses on leadership behaviors, (1) monitoring performance and (2) providing consequences. The model and its supporting research helps answer not only “what” leaders should do better to motivate employees, but also “how” they should go about doing so.

The Operant Model of Effective Supervision

The Operant Model of Effective Supervision consists of four major components: (1) the leaders' behaviors, (2) supervisory effectiveness, (3) the interaction process, and (4) moderators or the boundaries (see Figure 1). Leader behavior is comprised of three categories in the model which include providing antecedents, monitoring performance, and providing consequences. This is referred to as the antecedent-monitor-consequence (AMC) sequence.

Providing antecedents is defined as instructing, reminding, or conveying an expectation of performance (Komaki, 1998). When a leader gives an inspirational speech, delivers instructions, provide rules, reminders, training, or policy statements, or convey their expectations of performance, they are delivering performance antecedents. An example in a hospital might be a nurse manager saying, "Remember that we all need to be checking on our patients at least once every hour" to a group of nurses at the beginning of a shift. This would be considered an antecedent as the reminder precedes the behavior of interest. Another example of a leader providing an antecedent might be a leader saying "You should add the address and letter head in color to each page" after looking at her assistants memo she just wrote. In this case, the leader is providing instructions and conveying an expectation of performance.

Providing consequences, defined as communicating an evaluation or indicating knowledge of another's performance, was chosen as a component because of its similarity to a behavior in operant conditioning theory (Komaki, 1998). A general example of providing a consequence could be someone saying "Thank you so much!" after having the door held open for them. Providing consequences is further broken down by sign including positive, negative, and neutral. Positive consequences are when a leader expresses a favorable evaluation or approval of an employee's performance and the goal is to maintain or increase that performance in the future. For example, a leader saying "You saved us from potentially harming our customer" would be considered a positive consequence. Negative consequences are when a leader expresses disapproval or doubt about an employee's performance and the goal is to decrease or eliminate that specific behavior. For example, after walking by a nurse getting ready to deliver medication to her patient, a nurse manager saying "No, you've got the wrong dosage!" Neutral consequences are defined as expressing neither approval nor disapproval of an employee's actions. An example might be a manager asking a nurse "Have you asked this patient if they need to use the bathroom?" Following the nurse's answer, the manager might nod and say "Hmm." In this example, the manager has seen what the employee has done, but the evaluation would not be apparent.

Monitoring performance, defined as collection or gathering of information about followers' performance is the second major behavior in the Operant Model. Collection of information on behaviors is part of a three-step process used in operant motivational programs. For example, in a program to improve patient safety in a hospital, the operant approach would be to pinpoint the desired safety behaviors after reviewing the previous incidents, collect information by observing the actual safety behavior of workers (i.e., monitor performance), and then provide feedback (i.e., provide consequences) by sharing the information with the appropriate individuals. For these reasons, monitoring is a central component to the Operant Model. It is thought that managers who monitor will be more likely to gather accurate and timely information to provide meaningful consequences (Jensen & Komaki, 1993; Komaki, 1986, 1998). The model posits that the amount of time spent monitoring and providing consequences distinguish between effective and marginally effective leaders (Komaki, 1998). Simply stated, leaders judged to be effective are expected to monitor and provide consequences much more than their less effective peers.

The second component of the model, supervisory effectiveness, focuses on both follower performance and follower attitudes. To be considered an effective leader, they must have followers who perform well and have a positive attitude about their supervision (Komaki, 1998). A leader who has a team of followers that

perform at a high level but intensely dislike that leader would not be an example of effective supervision. Both performance and attitudes are considered complimentary components.

The third component of the model focuses on the interaction process (Komaki, 1998). The interaction process is situated between leaders' behavior and their effectiveness and focuses on the pattern of exchanges that occur when leaders and followers converse with one another. The patterns that facilitate effectiveness are those that are easy, frequent, and lively give-and-take between leaders and followers. The substance of the conversation makes a difference. The best exchange is one that is relevant to the task, and involves both parties discussing what each will do to get the job done. These are called performance-related discussions.

The fourth and final component of the model is the moderators. The leadership behaviors associated with the model are assumed to be most effective under specific circumstances. The five moderators identified help us understand when the model is likely to work best. The moderators consist of: (1) the characteristics of the followers, (2) the characteristics of the leader, (3) the resources of the organization, (4) the stage of the motivational process, and (5) the type of task (Komaki, 1998). Only when the level of these moderators is sufficient will the model work.

The first moderator focuses on the characteristics of the follower (Komaki, 1998). Only when the follower knows a job's technical aspects, possess the necessary skills, and have the potential to learn what is needed will the relationship between leader behavior and effectiveness be sustained. In addition, when followers are dissatisfied with their job because of duties, salary, or peers, they are unlikely to maintain high levels of performance or have a positive attitude regardless of the leaders' behavior in the model.

Just as knowledge, skills, and abilities are important, so are they for leaders. The second moderator concerns the characteristics of the leader (Komaki, 1998). Leaders must have detailed information about the structure they work within and the people within their work environment. They also require sufficient analytical skills. After monitoring, a leader must be able to make sense of the information they have obtained, use the data collected to generate hypotheses about what is going on, and then put what they have learned into action. Without these skills, they will be unable to correctly utilize the information gathered from monitoring.

The third moderator of the model takes into account the resources an organization has available (Komaki, 1998). No matter how well a leader engages in AMC sequences with employees, performance will not be sufficient if the equipment required to do the job is not available. When leaders find their budget

being chipped away year after year and are being asked to do more with less, the model is unlikely to provide much support.

The model was specifically designed to impact behavior at a certain stage of the motivational process. The stage in which the model will be most effective is the maintenance of behavior stage. To help clarify how the maintenance stage differs from other stages, a Taxonomy of Motivational Stages (TMS) was developed (Komaki, 1998). The model classifies the purpose of leader behavior into three categories: (1) to initiate, (2) to direct, and (3) to maintain. Initiating consists of a leader attempting to get employees doing something they are not currently doing. Directing occurs when a leader guides employees to learn a particular set of skills. In maintaining behavior, the focus shifts from behaviors that are not known to those that employees are already carrying out on a regular basis. The emphasis is therefore on sustaining these behaviors over time. It is important to note that only when a leader is attempting to maintain behavior is the OMES an appropriate model to follow.

The final moderator is the type of task (Komaki, 1998). Only when execution of a task is of primary concern is the OMES appropriate. When a leader is primarily concerned with ensuring that a certain task gets accomplished, the OMES will be of immense value. Other tasks, such as decision making or delegation are not what the model is designed to support. It is not that these other

tasks are less important, but we must acknowledge the type of task the model was designed for. By understanding these five moderators, it becomes clear when the model can and cannot be successfully applied.

Research on the Operant Model

A sizeable amount of research exists showing relationships between behaviors in the Operant Model and leadership effectiveness (Brewer, Wilson, & Beck, 1994; Jensen & Komaki, 1993; Komaki, 1986; Komaki, Desselles, & Bowman, 1989; Komaki, Hyttinen, & Immonen, 1991; Komaki, Reynard Minnich, & Wallace, 1997). Eight field studies have been conducted in which the relationship between leadership effectiveness and the behaviors in the Operant Model were assessed. In all eight studies, the time leaders spent monitoring or providing consequences, or both, were related to the specified criteria for leadership effectiveness. Leader behavior was measured primarily via an observational tool, known as the Operant Supervisory Taxonomy and Index (OSTI), in which observers' watch leaders behave in their natural environments and record what they say and do.

Komaki, Zlotnick, and Jensen (1986) conducted a study to evaluate the use of the OSTI as a tool to observe and measure leader behavior in real time. The OSTI was used to observe seven theater managers over a 5-week period and 20 bank managers over a 12-week period. Researchers used the tool to observe

managers for random 30-minute periods. Observers' collected data on time spent engaging in a total of seven categories of leadership behavior during each 30-minute observation session. The first category is performance consequences (1) which was defined as, "communicates an evaluation of or indicates knowledge of another's performance, where the indication can range from highly evaluative to neutral." The second category, performance monitors (2), was defined as, "collects information about a follower's performance." Performance antecedents (3) was a third category defined as, "instructs, reminds, or conveys an expectation of performance." Own performance (4) was a category of leadership behavior defined as "refers to leader's own performance." The category of work related (5) was described as "refers to work issues but not to worker performance." Nonwork related (6) was defined as "does not refer to work issues" and solitary (7) was defined as "does not interact with others."

Categories were further broken down into subcategories in certain cases. For example, performance consequences was also broken down into subcategories in terms of delivery (i.e., direct or indirect) and evaluation (i.e., positive, negative, or neutral). Performance monitors was also broken down into the subcategories of work sampling, product sampling, self-report, and secondary source. See Komaki (1998) for a complete list of definitions. Approximately 189 30-minute

observations were conducted for the 7 theater managers and 440 30-minute observations were conducted for the 20 bank managers.

Results from the study found that the OSTI was feasible, acceptably reliable, and sensitive to differences in behavior among managers. The final analysis indicated that 20 30-minute observations would provide representative information on a given manager's behavior. In other words, by using the OSTI, researchers can effectively observe and collect data to gather information on how leaders spend their time. This groundbreaking study helped address one of the ongoing challenges in conducting observational research with leaders of reliably observing leaders behavior in action.

Shortly after the validation of the OSTI as a reliable tool to measure leadership behavior, a study was conducted using the new tool to assess leadership effectiveness. Komaki (1986) used the OSTI to observe two separate groups of managers in a medical insurance firm. The first group consisted of 12 managers who were ranked by their superiors in the top 28% in terms of motivating others. The second group consisted of 12 managers who ranked in the bottom 28% in terms of motivating others. A total of 465 30-minute observations were made over a 7-month period. Results indicated that effective managers spent significantly more time collecting performance information than their less effective peers ($t = 2.59, p = .009$). Specifically, effective managers spent more time work sampling, inspecting

the work itself, or watching employees conduct work. One limitation to the study is the enormous amount of time required to collect reliable information on leadership behavior. Business leaders are unlikely to invest the money and resources required to conduct this many observations.

In another study, Komaki, Desselles, and Bowman (1989) used an expanded version of the OSTI which included “team coordination” as an additional subcategory. Team coordination occurred when a leader referred to two or more team members’ actions which intersected. The tool was used to observe 19 skippers during a sailboat regatta. Data were collected during both the preparation phase of each race and during the race itself. The preparation phase began the moment the skipper first set foot on the boat until the 3-minute warning whistle. The race began at the 3-minute warning whistle and ended as soon as the boat crossed the finish line. In addition to leader behavior data collected during the race, head coaches ranked (1 = highest) and rated (A through F) each of the participating skippers in terms of crew handling.

Results of the study showed that series standings were significantly correlated with performance consequences ($r = -.47$, $p < .05$) and monitors ($r = -.51$, $p < .05$) during the races themselves. In addition, observations of skipper performance was correlated with coach ratings and ranking of crew handling. Results showed that providing consequences during the race was significantly

correlated with ratings and ranking of crew handling ($r = -.60, p < .05$). No relationship was found for monitoring ($r = -.42, p > .05$). These findings are important as they demonstrated a significant relationship between leaders monitoring and providing consequences and an objective outcome. However, the time required to observe each skipper was 94 minutes. This was actually considered a limitation in the study as more observations are typically needed to gather accurate data.

The next study looked at two sets of leaders in Finland. Komaki, Hyttinen, and Immonen (1991) assessed 31 managers in the construction industry and 16 supervisors in a government agency. To measure leadership effectiveness for managers in the construction industry, construction crews rated their manager on a five-point scale in terms of how well they met, “the demands of the job.” Crews rated on a 5-point scale from 1 (very bad) to 5 (very good). Researchers found that effective construction managers spent significantly more time providing feedback compared to the less effective managers ($t = 1.7, p = .05$).

For the government agencies, researchers measured leadership effectiveness with satisfaction and well-being questionnaires. To measure satisfaction, the Job Diagnostic Survey (Hackman & Oldham, 1975) was used which asks questions related to general satisfaction, growth satisfaction, satisfaction with job security, and satisfaction with supervision. To measure well-being, researchers used the

Eigenzustandsskala (EZ-Scale) (Nitsch, 1976) which assesses a person's feelings toward her job at a certain moment. Employees were asked to rate themselves on specific questions related to self-confidence, sociability, and state of mind. Ratings ranged from one to six with six being the most positive. Results showed that supervisors who provided more monitors and consequences had workers with greater well-being. In addition, these supervisors also had employees who were more satisfied with job security, particularly if the manager monitored more often ($r = .82$ to $.87$). A positive relationship was also found between managers who provided more consequences and workers' reported positive feelings about their mental effort, sociability, self-confidence, state of mind, and tolerance ($r = .73$ to $.89$). These results are encouraging, however, it is unclear the extent to which satisfaction and well-being questionnaires are related to performance in this particular setting or others.

A study conducted in a daily newspaper operation used an extreme groups design to compare the top ($n = 8$) and bottom ($n = 8$) quarter of managers in terms of leadership effectiveness. Measures of leadership effectiveness were ratings from superiors within the organization. Approximately 30 20-minute observations were conducted using the OSTI for each manager to measure frequency of monitoring and providing consequences. Results from the study found that top-rated managers spent more time providing consequences than their poorly-rated counterparts ($t =$

3.42, $p < .05$). In other words, effective managers were more likely to provide positive, neutral, and negative consequences to their staff. Another interesting finding from the study was that bottom-rated managers ($M = 13.8$ percent) were more likely to be silent while engaged in a face-to-face exchange than top-rated managers (10.9 percent), $p < .05$ (Jensen & Komaki, 1993). Unfortunately, ratings from superiors tend to be very subjective and may be a limited criteria in the given study. In addition, approximately 160 hours of observations were required to gather meaningful data on leader behavior. This amount of time is impractical and unlikely to be adopted by organizations today.

Brewer, Wilson, and Beck (1994) examined Komaki's operant model of effective supervision with 20 police patrol sergeants within an Australian police force. A measure of leadership effectiveness was provided by superior officer's ratings of the performance of each patrol sergeant's team of subordinates. The OSTI was used to conduct 20 30-minute observations for each of the 20 police patrol sergeants over a four month period of time. Results from the study found that supervisor's use of performance monitoring was significantly related to higher team performance ($r = .40$, $p < .05$). Additionally, a significant relationship was found between supervisor's providing neutral consequences and team performance ($r = .51$, $p < .05$). These findings help provide support for the generalizability of the model given the environment of study. That being said, over 200 hours of

observation were required to observe all 20 police patrol sergeants. One can assume that a typical organization is unlikely to deploy such methods to measure leadership behavior for selection or training purposes.

Methot, Williams, Cummings, and Bradshaw (1996) evaluated the impact of a supervisor training program on manager, supervisor, and staff behavior in a residential facility. One manager and four supervisors completed a three-hour training session consisting of a didactic presentation covering goal setting and feedback in individual sessions with staff. Following the presentation, trainees watched a video describing the form and content of performance feedback in similar settings. The video included a demonstration of objective performance monitoring and contingent feedback to subordinates. The OSTI was used to collect data on the use of objective measures and contingent consequences by the manager and supervisors. Data were also collected on whether or not changes in supervisor behavior results in changes in the direct care staffs use of contingent consequences for client performance and changes in client behaviors. An increase in the use of contingent performance consequences was observed for the manager and all four supervisors. In addition, an increase in the use of contingent consequences was seen in six of the seven direct care workers. Desired increases in client target behaviors were seen for 8 of 13 clients and desired decreases in client target behaviors were observed for 9 of 16 clients. These findings are important as they suggest that by

monitoring and providing contingent performance consequences, measurable improvements in direct staff performance and client outcomes might be expected.

Komaki, Reynard Minnich, Lee, and Wallace (1997) assessed 28 leaders of teams on sailboats in a fleet racing competition. The primary measure of leader effectiveness was the time took to hoist the sail. Hoisting of the sail is a maneuver that requires five crew members to simultaneously carry out a minimum of six precisely timed steps. Faster sail hoisting is preferable in this context. Video cameras were used to capture behavior on the boats during each race. Results from the study found that leaders monitoring of performance was significantly associated with faster hoist times ($r = -.33, p < .05$). Leaders who monitored the equipment (e.g., looking up at the sail to see the impact of the crew's actions) were more likely to have faster sail hoists than leaders who did not monitor this way. These results are encouraging as they link the leadership behavior of monitoring to a performance outcome. The gap that remains in this particular study is a measure of team member perceptions. It might be the case that team members on the winning sailboat might not wish to sail under that skipper again. As Daniels and Daniels (2007) state, "When it comes to leadership, how you accomplish success is every bit as important as what you accomplish" (p. 3). Having a measure that captures follower attitude and perception and that is also associated with performance outcomes would be most ideal in this case.

One study sought to go beyond the operant model itself to study the process involved in performance monitors (Komaki & Citera, 1990). The study hypothesized that performance monitors set into motion a reciprocal and performance-related set of events between superiors and subordinates. The study deployed a randomized two-group design which included 60 manager-subordinate pairs. One group consisted of managers who monitored performance (e.g., sampled work) and the other group consisted of managers who provided antecedents (e.g., provided instructions). Researchers collected data on both superior and subordinate behavior during each interaction. Results from the study found that subordinates spent more time talking about their own performance in the monitor versus antecedent group ($t = 3.14, p < .05$). In addition, managers in the monitor group provided significantly more consequences than managers in the antecedent group ($t = 7.81, p < .05$). Managers in the monitor group were able to provide more superior consequences and stimulate discussions from subordinates regarding their own performance than managers in the antecedent group. This, in turn, set the stage for managers to provide consequences and continue monitoring subordinates' performance.

A sizeable amount of research has been conducted to validate the Operant Model of Effective Supervision. Additionally, the OSTI appears to be a valid and reliable tool practitioners and organizations can use to capture leadership behavior.

However, the amount of time and resources required to use the OSTI and gather meaningful information (approximately 10 hours per leader) is not well-suited to become an operational predictor used by leaders in organizations. Whether an organization is attempting to assess current leadership talent or assess the abilities of a potential candidate, measures of knowledge, skills, and abilities need to be of short duration.

Operant Supervisory In-Basket Assessment (OSIBA)

As a response to these practical concerns, an in-basket assessment was created to measure the behaviors in the Operant Model called the Operant Supervisory In-Basket Assessment (OSIBA) (Komaki, Newlin, & Desselles, 1990). An in-basket exercise is an assessment of an individuals' demonstrated skills that requires minimal time commitment. In-baskets are simulations in which respondents take on the role of a manager and respond to items that a manager might find in their "IN" bins. These items typically consist of fictional memos, phone messages, letters, emails, etc. Respondents are asked to respond to items, make decisions, or attempt to resolve issues created by the items (Brass & Oldham, 1976; Gill, 1979; Schippmann, Prien, & Katz, 1990). In general, in-basket exercises do not take more than a few hours for respondents to complete. Scoring of the in-basket can be cumbersome as scoring must be done by hand and respondents can answer in a seemingly infinite number of ways.

The OSIBA is a paper-and-pencil measure in which respondents take on a leadership role by handling items in their “in-baskets.” In the assessment, you are put in the role of the editor of a monthly food magazine company. A total of 16 in-basket items are on the assessment for leaders to respond to, ranging from telephone messages from clients complaining about poor service to reports from accounting in production costs. Respondents respond freely to whoever and in whatever way they see most appropriate. The in-basket exercise typically takes about one hour to complete. One study investigated the utility of the OSIBA as a replacement measure for the OSTI to save considerable time and effort of the leaders under study. Komaki et al. (1990) conducted a study with 12 computer managers. Each manager was observed using the OSTI, and scored for their responses on the OSIBA. Results showed that scores for monitoring and providing consequences as scored on the OSIBA were significantly correlated with the amount of time spent on the same observed categories using the OSTI (monitors, $r = .57, p < .05$; consequences, $r = .60, p < .05$). Meaning, how a leader responds to items on the OSIBA is reflective of how a leader would respond if observed in the same situation.

Reynard Minnich (2007) attempted to link the leadership measures on the OSIBA with supervisory effectiveness. Researchers asked 35 investment bankers to take a modified version of the OSIBA. The assessments were scored for the

behaviors of monitoring, providing consequences, and providing antecedents. In addition to the in-basket, ratings of motivational effectiveness and technical expertise were collected from each participants' supervisors, subordinates, and peers. Results from multiple correlations found that overall monitoring ($r = .30, p < .05$) and monitoring via work sampling ($r = .28, p < .05$), were significantly correlated with motivational effectiveness. Providing consequences that were considered "warranted" or "contingent" were positively and significantly correlated with motivational effectiveness ($r = .29, p < .05$). Providing positive consequences were significantly correlated with motivational effectiveness ($r = .36, p < .05$). Another measure scored under the providing consequences category was "thanking the bearer of bad news." Respondents earned points in this category when they recognized individuals or thanked them for bringing difficult information to their attention. Leaders scores for providing consequences in which they thanked the bearer of bad news was significantly related to their rating of motivational effectiveness ($r = .40, p < .01$).

A major limitation, thus far, of the studies mentioned that used the OSIBA is the measure of motivational effectiveness. Though it is encouraging to see such strong relationships between responses on the OSIBA and motivational effectiveness, it is unclear to what extent ratings of "motivational effectiveness" can be used as a measure of leadership success. Industrial-Organizational Psychology

researchers have established the inclusion of both employee/team performance and follower attitudes as common practice when defining leadership effectiveness (Komaki, 1998).

One study used the OSIBA to measure the extent to which leader's behavior changed following a training on the operant model. Komaki, Minnich, Grotto, Weinshank, and Kern (2011) conducted two experiments in which training was provided to seasoned managers on the Operant Model of Effective Supervision. In the first experiment, trainees consisted of twelve lower to upper level managers in a family-owned merchandising agency. Participants were randomly assigned to a treatment and control group. Approximately seven participants were in the treatment group, and five participants in the control group. Training was provided to the treatment group by one of the researchers. The training took approximately 8 hours and was broken down into 2-hour sessions held weekly across 4 weeks. The training covered the Operant Model's behaviors of monitoring and providing consequences. One week after the completion of the training in experiment 1, both the control and treatment groups took the OSIBA. Results from the OSIBA showed that trainees demonstrated a substantial increase in monitoring, providing positive and neutral consequences, and thanking the bearer of bad news.

In experiment 2, participants consisted of 63 managers from the EMS Operations unit of a major metropolitan fire department. Similar to experiment 1,

participants were randomly assigned to a control and treatment group. The treatment group consisted of 32 managers and the control group consisted of 31 managers. The treatment group went through a single 5-hour training session on the Operant Model. For experiment 2, the control group took the OSIBA before the training occurred and the treatment group took the OSIBA after training. Results from the OSIBA showed that members from the treatment group significantly increased monitoring, providing consequences, and thanking the bearer of bad news. For example, members in the treatment group thanked the bearer of bad news 11% more often than members of the control group ($p = .003$).

Of the eight studies thus far that assessed the relationship between the OMES and varying measures of motivational effectiveness, seven of the eight found a significant relationship between monitoring performance and motivational effectiveness (Brewer, Wilson, & Beck, 1994; Komaki, 1986; Komaki, Desselles, & Bowman, 1989; Komaki, Hyttinen, & Immonen, 1991; Komaki, Reynard Minnich, Lee, & Wallace, 1997; Methot, Williams, Cummings, & Bradshaw, 1996; & Minnich, 2007). These studies provide support for specific leadership behaviors (i.e., monitoring and providing consequences) being predictive of leadership effectiveness. If this is true, any measure that is truly reflective of leadership effectiveness should also be significantly related to performance monitors and providing consequences. Therefore, **hypothesis 1a** of the current study was that

monitoring would be positively and significantly related to employee engagement scores.

In addition, six of the eight studies found significant relationships between the supervisory behavior of providing consequences and varying measures of motivational effectiveness (Brewer, Wilson, & Beck, 1994; Jensen & Komaki, 1993; Komaki, Desselles, & Bowman, 1989; Komaki, Hyttinen, & Immonen, 1991; Methot, Williams, Cummings, & Bradshaw, 1996; & Minnich, 2007). Thus, **hypothesis 1b** of the current study was that **providing consequences would be positively and significantly related to employee engagement scores.**

Previous support for providing antecedents is less robust. Reynard and Komaki (1995) found a significant link between antecedents provided before performance monitors and a measure of motivational effectiveness. Minnich (2007) did find a significant positive relationship between bank supervisors providing antecedents and subordinates rating of their technical skill. One study found that providing antecedents functioned as a suppressor variable for the other behaviors in the OMES (Jensen & Komaki, 1993). As such, **hypothesis 1c** of the current study was that **providing antecedents will be negatively and significantly related to employee engagement scores.**

As discussed, the OMES suggests that monitoring and providing consequences interact. An additional proposition of the model suggests that a

leaders' monitoring and providing consequences behaviors “work together” for superior leadership effectiveness (Komaki, 1998). The nature of the interaction may be such that when monitoring is high, the relationship between providing consequences and leadership effectiveness is positive, but when monitoring is low, providing consequences will not be related to leadership effectiveness. With this theoretical rationale, the interaction of monitoring and providing consequences may be useful in predicting leadership effectiveness.

In the current study, a hierarchical multiple regression including monitoring, providing consequences, and the multiplicative combination of the two, was conducted. If an interaction exists, the total amount of variance explained when the multiplicative term is part of the regression equation will be significantly more than that explained when the regression equation consists of the component parts of the interaction alone. Thus, **hypothesis 2** of the current study was that **the interaction of monitoring and providing consequences will add significantly to the prediction of employee engagement, over and above monitoring and providing consequences alone.**

Not only has the OSIBA been validated to reliably predict how a leader will spend their time, measures from the in-basket have also been reliably correlated with a measure of leadership effectiveness. In addition, the OSIBA has been used to identify behaviors that can be improved with a training program. One of the

shortcomings of the research to date is the criteria used to determine leadership effectiveness. The previous studies have mostly used ratings of motivational effectiveness from high level leaders, peers, and subordinates. The decision to use these rating measures was sound as most times, a single outcome is not available for which all leaders in a given study can consistently be judged.

Ratings like these do not meet the criteria for supervisory effectiveness as originally defined in the Operant Model of Effective Supervision. A true measure of leadership effectiveness would capture the performance of employees and their attitudes toward their leader. Therefore, team performance and team perceptions are both important measures to consider when studying leadership behavior. Finding a single measure that captures follower attitudes and is predictive of team performance would be most ideal. One proxy to performance outcomes and employee perception that might serve as a more robust leadership effectiveness criterion is a measure of employee engagement.

Employee Engagement

Employee engagement is a construct that has become the latest “buzz” word in management (Ludwig & Frazier, 2012). The first major article on engagement to appear in the Industrial-Organizational Psychology literature was Kahn’s (1990) article on personal engagement and disengagement. Since Kahn’s article, much research has been conducted on the concept. Multiple definitions have emerged in

the literature regarding employee engagement. Kahn (1990) defined employee engagement as, “the harnessing of organization members’ selves to their work roles; in engagement, people employ and express themselves physically, cognitively, and emotionally during role performances” (p. 694). Christian, Garza, and Slaughter (2011) defined engagement as a construct that, “involves a holistic investment of the entire self in terms of cognitive, emotional, and physical energies” (p. 97). Maslach and Leiter (2008) defined engagement as, “an energetic state of involvement with personally fulfilling activities that enhance one’s sense of professional efficacy” (p. 498). Schaufeli, Salanova, Gonzalez-Roma, and Bakker (2002) describe engagement as “a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption” (p. 74).

Gallup, a leading company in employee engagement surveying defines engagement as, “employees who are involved in, enthusiastic about and committed to their work and workplace” (Gallup, 2017). Other words used to define employee engagement such as “vigor” or “energetic” as Ludwig and Frazier (2012) put it, “may be simply a matter of managing contingencies through the behavioral systems that include operational and managerial processes” (p. 75). Over time, a common theme has emerged regarding research findings on employee engagement. That is, employee engagement is key to the success and competitiveness of an organization (Saks & Gruman, 2014).

Although multiple definitions exist, Daniels (2009) defines employee engagement simply as, “a non-specific non-scientific term used to describe the amount of positive reinforcement available in a workplace for value-added behavior” (p. 7). To take this one step further, it is likely that when an employee is engaged, they are emitting specific observable behaviors. Multiple of the previous definitions use constructs such as “dedication” or “committed” or “enthusiastic” to describe employee engagement. Although these terms have not been previously defined from a behavioral view, “dedication” or “commitment” might include behaviors such as (1) showing up to work on time, (2) completing job duties not in their job description, or (3) volunteering to be part of committees or teams. Daniels and Bailey (2014) refer to behaviors of this nature as “discretionary effort.” These are just a few of many behaviors that likely make up an “engaged” employee. Organizations are complex environments and trying to measure every individual behavior that makes up an “engaged” employee would be nearly impossible today. Having a simple tool to measure the likelihood these behaviors are occurring or not in a large group of employees can be both valuable and powerful information for business leaders to acquire.

One of the reasons employee engagement has received so much attention is that employee engagement scores are closely associated with employee and organizational outcomes. For example, engagement has been positively linked with

job performance and organizational citizenship behavior (Bakker & Bal, 2010; Rich, LePine, & Crawford, 2010; Saks, 2006) and negatively related to turnover intentions (Saks, 2006; Schaufeli & Bakker, 2004). Bakker and Bal (2010) measured weekly work engagement with 54 Dutch teachers over 5 consecutive weeks using a revised version of the Utrecht Work Engagement Scale. The instrument consists of 9-items which measure three domains including vigor, dedication, and absorption. Job resources were also measured weekly which included autonomy, social support, exchanges with supervisor, and opportunities for development. Results from the study found weekly levels of autonomy, exchanges with supervisor, and opportunities for development had a significantly positive relationship with weekly levels of employee engagement scores. In addition, weekly levels of employee engagement scores were positively related to job performance. It is likely that specific leadership behaviors occur on a day-to-day basis, which lead to environments of autonomy and opportunities for development.

Saks (2006) conducted a study with 102 employees across various organizations that examined a model of the antecedents and consequences of employee engagement based on social exchange theory. All participants completed a survey that measured employee engagement and a host of antecedents and consequences to employee engagement. Antecedents to engagement were factors

thought to predict employee engagement. Antecedents that were measured included job characteristics, perceived organizational support, perceived supervisor support, reward and recognition, procedural justice, and distributive justice. Consequences of employee engagement included the possible results from employee engagement on an organization. Consequences that were measured included job satisfaction, organizational commitment, intentions to quit, and organizational citizenship behavior. Results from the study found that a significant relationship exists between employee engagement and job satisfaction, organizational commitment, intentions to quit, and organizational citizenship behavior. Harter, Schmidt, and Hayes (2002) analyzed data across 7,939 business units in 36 companies and found that employee engagement was related to specific business-unit outcomes such as customer satisfaction, productivity, profitability, turnover, and safety. In another study, Macey, Schneider, Barbera, and Young (2009) found that in 69 different firms across multiple industries, the top 25% on an engagement index had greater return on assets (ROA), profitability, and more than double the shareholder value compared to the bottom 25%.

The Gallup Q12 measure of employee engagement is a commonly used 12-item scale that measures employee engagement ("Gallup Employee Engagement Center"). This instrument is one of the industry's leading tools used to measure employee engagement in organizations throughout the world today. Since the

creation of the survey in the 1990's, more than 25 million employees, from 189 countries, speaking 69 different languages have completed the survey. The tool boasts over 30 years of both qualitative and quantitative research (Harter, Schmidt, Agrawal, Plowman, & Blue, 2016). The company has focused primarily on the inclusion of items they deemed "actionable" at the supervisor or manager level. The tool addresses items such as role clarity, resources, receiving feedback, fit between abilities and requirements, and feeling appreciated. Employees are asked to respond to each of the 12 statements using six Likert-style response options from 5 (strongly agree) to 1 (strongly disagree) with a sixth option of "don't know/does not apply" which goes unscored (Harter, Schmidt, Agrawal, Plowman, & Blue, 2016).

A recent Gallup meta-analysis reviewed 339 research studies across 230 organizations in 49 industries, with employees in 73 different countries. A total of 82,248 business units were assessed across these organizations that included a total of 1,882,131 employees. Relationships between employee engagement as measured by the Gallup Q12 survey and business outcomes were calculated. The business outcomes included were customer loyalty/engagement, profitability, productivity, turnover, safety incidents, shrinkage, absenteeism, patient safety incidents and quality (defects). Results of the analysis found that employee engagement is significantly correlated to each of the nine outcomes studied. Business-units at the 99th percentile for employee engagement had four times the success rate of those at

the first percentile. In addition, median differences between top-quartile and bottom-quartile units were 10% in customer ratings, 21% in profitability, 20% in sales production, 17% in production records, 24% in turnover (high-turnover organizations), 59% in turnover (low-turnover organizations), 70% in safety incidents, 28% in shrinkage, 41% in absenteeism, 58% in patient safety incidents and 40% in defects (Harter, Schmidt, Agrawal, Plowman, & Blue, 2016). Clearly, producing a highly “engaged” workforce is a sound strategy for any business looking to improve their outcomes. The question then becomes, how can leaders improve employee engagement? Being that employee engagement captures both employee attitudes and is highly correlated with employee performance, it may serve as a helpful measure of leadership effectiveness when evaluating the OSIBA.

Creating Employee Engagement

Ludwig and Frazier (2012) suggest that management behavior mediates the relationship between employee engagement and organizational outcomes. It is estimated that approximately 70% of the variance in employee engagement scores across business units can be accounted for by managers (Beck & Harter, 2015). One question on the Gallup survey asks employees whether or not someone has talked with them about their progress at work within the last six months. Another question asks whether or not the employee has received recognition within the last seven days. Employees may be more likely to respond favorably to these questions

if they work for a leader who consistently monitors and provides performance based consequences. Another question asks employees if they feel their opinion counts at work. Consistent monitoring of performance and providing consequences may create the conditions for employees to express their opinions more frequently. The leadership behaviors measured in the Operant Model are likely to create the conditions for employees to respond positively on each of the questions asked on the employee engagement survey.

In general, job resources such as autonomy or job control, coaching, feedback, and opportunities for development have been found to be positively related to employee engagement (Bakker, Albrecht, & Leiter, 2011; Bakker, Hakanen, Demerouti, & Xanthopoulou, 2007; Mauno, Kinnunen, & Ruokolainen, 2007; Xanthopoulou, Bakker, Demerouti, & Schaufeli, 2009a). Additionally, Schaufeli, Bakker and Van Rhenen (2009) conducted a two-wave longitudinal study with a 1-year interval of managers and executives in a Dutch Telecom company. Over 400 managers received a paper questionnaire at the start of the study and one year later. The survey measured job resources (social support, autonomy, opportunities to learn and develop, and performance feedback), burnout, work engagement, and sickness absence. Results after one year suggest that when job demands increase, and job resources decrease, future burnout scores increase. Moreover, as job resources increase (e.g., social support, performance feedback,

etc.), work engagement increases. Therefore, **hypothesis 3** was that **providing antecedents for motivation will improve the prediction of employee engagement over and above monitoring and providing consequences**. That is, more variability in employee engagement scores will be predicted by providing antecedents, monitors, and consequences than by providing monitors and consequences alone.

These findings are important as they help leaders understand “what” must happen to get the most out of their employees (i.e., social support, performance feedback, etc.). However, a gap still exists in helping leaders understand “how” to execute these essential concepts such as increasing social support and providing performance feedback. Knowing that performance feedback is associated with employee engagement is helpful, but begs the question “what does that look like in action?” Whether the performance feedback is positive or negative, frequent or infrequent can only be determined by observing leaders with high employee engagement scores. Bakker et al. (2011) suggest that future research should examine whether, “leaders also influence followers’ work environment, and indirectly their work engagement” (p. 14). Determining which leadership behaviors are most critical to driving employee engagement is a question yet to be fully answered by the research.

The current study sought to use a measure of leadership effectiveness that is more robust and readily recognized by organizations and research. A plethora of research exists correlating employee engagement with organizational outcomes. These findings provide us with a universal outcome to use as a measure of leadership effectiveness. Thus, the current study sought to correlate leadership behavior as measured by the OSIBA with employee engagement scores as measured by the Gallup Q12 survey. It was anticipated that employee engagement would serve as a reliable measure for leadership effectiveness.

Method

Participants and Setting

The current study took place with 47 individuals holding a leadership role in a hospital system in the southeastern United States. The system was made up of four acute-care hospitals ranging in size from 100 beds to 500 beds. Each of the four hospitals offered a variety of services from emergency services to inpatient surgical procedures. Participants consisted of leaders of major operational departments in the hospital system including nursing, radiology, environmental services, transport, pharmacy, etc. Participants were also required to have a minimum of five direct report responses on the employee engagement survey. This requirement was necessary to be able to access the individual employee engagement score for each leader, which served as the criterion variable. The

surveying agency withholds all employee engagement scores with fewer than five total employee responses. The study took place in parallel with the yearly administration of the employee engagement survey to all associates in the organization.

Only leaders who were in their position since the administration of the prior employee engagement survey were included in the study. The organization had conducted this survey yearly for many years. The surveys were assessed from the previous administration to develop the sample pool for the current study. Leaders who met two criteria were selected as potential participants: (1) had at least five responses on the last survey administered by the organization, and (2) were still in a leadership position at the time of the study.

A total of 65 leaders met the criteria and were extended the opportunity to participate in the study. Two of these leaders were unable to be reached and never had the opportunity to participate. A total of 63 leaders were provided an assessment and given an equal and voluntary opportunity to participate. A total of 49 leaders elected to participate resulting in a 77.78% participation rate. Two of the 49 leaders did not end up with at least five employee responses resulting in a suppressed employee engagement score (i.e., employee engagement score was unavailable). As a result, their responses could not be included. Therefore, the final participation total was 47 leaders.

Additional demographic information was collected during the study including age, gender, years of education, number of years in current position, number of year in the organization, years of supervisory experience (including other organizations), and days of supervisory training while employed in this organization. The majority of the sample was female (78.7%, $n = 47$). Most participants reported having an undergraduate degree (90.1%, $n = 44$). Fewer participants reported having a graduate degree (33.3%, $n = 24$). On average, participants were 48.1 years of age, had 16.3 years of education, 11.1 years tenure at the organization, 4 years in their current position, 10.1 years of supervisory experience, and 13.2 days of supervisory training. See Table 2 for information about demographics of the sample.

An additional question asking the participant to rate their level of motivation to perform well on the assessment was included at the end of the assessment (Penk & Schipolowski, 2015). Participants' responded on a Likert-scale ranging from 1 (not at all motivated) to 4 (highly motivated). Responses ranged from one to four with the majority of respondents reporting to be highly motivated to perform well ($M = 3.47$, $n = 47$).

Design

The current study conducted a hypothesis test and analyzed the relationship between leadership behavior and employee engagement scores.

Operant Supervisor In-basket Assessment (OSIBA)

Participants were asked to take an in-basket assessment called the Operant Supervisory In-basket Assessment (OSIBA) to measure leadership behavior. The OSIBA was originally designed by Komaki (1998) as an alternative to the Operant Supervisory Taxonomy Index (OSTI). The OSIBA is an exercise where participants took on the role of a publisher of a monthly food magazine company. Participants were then asked to provide written responses to items in their “in-baskets” however they deemed appropriate. Items found in the in-basket came in the form of memos, written phone messages, and e-mail messages from employees and department directors and participants were asked to write responses to characters depicted in the scenarios.

OSIBA Categories and Scoring

The OSIBA was scored for three different categories of supervisory behaviors including monitoring, providing consequences, and providing antecedents. Monitoring was divided by method (via work sampling and self-report); antecedents by type (used for motivation and tacking-on traditional antecedents); and consequences by sign (negative and positive/neutral). Positive consequences were further broken down into five subcategories including (1) providing simple, short acknowledgements for standard reports; (2) giving traditional positives for effort or accomplishments; (3) thanking the bearer of bad

news; (4) acknowledging someone out of the contact loop; and (5) broadcasting congratulations to multiple parties. Two neutral consequences were included: (1) letting the sender/other(s) know of actions taken; and (2) relaying a problem nonjudgmentally. Each category and subcategory listed had an operational definition. For example, a work sample was defined as “indicating interest in observing workers in action and/or examining work products” (Komaki, Minnich, Grotto, Weinshank, & Kern, 2011). A full list of definitions can be found in Table 1. Permission to include a full copy of the assessment and scoring was not granted by the proprietary owner of the tool.

Scoring

Responses to each item on the assessment were scored category by category by trained raters. Points were allocated depending on the quality of monitors, antecedents, and consequences. For example, based on the operant model, a monitor via work sample was considered more effective than a monitor via self-report. Therefore, a work sample was worth more points than a self-report. As many as six points were awarded for a monitor via work sample. A response to one item on the assessment had the potential to include multiple categories. For example, a response could have included a monitor and a positive/neutral consequence. See Table 3 for the total points possible for each category and subcategory by item.

Overall scores were calculated for each category by summing the subcategory scores. For example, monitoring overall consisted of the sum of work sample and self-report points. The total score was divided by the total points possible for that category and multiplied by 100 to yield a percentage score. If a participant received 17 points for monitoring via work sampling and 11 points for self-report, the overall raw score for monitoring was 28 points. A total of 66 points for monitors were possible to obtain on the OSIBA, so the participant's monitoring score was calculated as 28 divided by 66 and multiplied by 100, for a score of 42%. The same methodology was used to obtain a 'providing antecedents' and 'providing consequences' score for every participant.

Interrater Agreement

Scoring was conducted independently by the primary researcher and two secondary raters. The primary researcher scored every response from all participants. The secondary raters each independently scored 50% of the total number of assessments completed. Therefore, all 47 assessments were scored independently by the primary and at least one secondary rater.

Interrater agreement was calculated as the number of agreements divided by the total number of points scored, multiplied by 100. If the primary rater awarded a participant 30 points for monitors and the secondary rater awarded that same participant 28 points, then interrater agreement was 28 divided by 30, multiplied by

100 giving an agreement score of 93.3%. Overall agreement was calculated for monitors, providing consequences, and providing antecedents. If agreement did not equal a minimum of 80%, each researcher would re-score until 80% agreement was reached. All major categories received agreement scores above 80% and no rescoring was necessary. All disagreements were then reconciled on an individual basis by all three members of the scoring team. Final reconciled scores were used for the study.

Criterion: Employee Engagement

An employee engagement survey administered once every twelve months was used as the measure of effectiveness for each leader. The survey, administered by Gallup, consists of 12 questions that employees respond to on a Likert scale. Questions range from asking employees about being recognized in the last seven days to feeling as though their opinion counts in the workplace. Every leader in the organization who has direct reports gets a score every 12 months. The most recent score was used which was received the same month the OSIBA was administered.

The sum of responses was averaged for each question resulting in a mean score for that question and a percentile rank. The percentile rank is where that individual leaders score falls compared to all other scores in the Gallup database across multiple industries. Finally, a score is calculated for all 12 questions to get an overall engagement score. This score is also provided as a grand mean score and

a percentile rank. Individual percentile rank scores were used as the measure for employee engagement in the current study.

Procedure

To begin the project, the Chief Human Resources Officer of the organization sent an email explaining the evaluation of the in-basket assessment as an organizational initiative. The email explained that the individuals receiving the assessment had been randomly selected to participate in an internal voluntary study to validate the tool. The email also explained that they would be receiving a paper version of the assessment by the experimenter and given 3-weeks to participate.

Within two days of the email being sent, the experimenter hand delivered a package to each of the selected participants. The package included the following: (a) a title page, (b) a letter giving the deadline for completing the exercise and instructions for returning it, (c) a license agreement, (d) instructions for responding to the in-basket items and describing the fictitious situation, (e) the 16 in-basket items followed by space to respond, and (f) a demographic information page. After the first week, the experimenter reached out to each individual in person to remind them that they had two weeks left if they wished to participate. A final reminder was sent out via email to all potential participants when only two days remained to participate.

Analysis

To test hypothesis 1a, 1b, 1c, zero-order Pearson correlations were computed between categories of supervisory behavior assessed in the in-basket and employee engagement scores. An additional hierarchical multiple regression was conducted to control for the sample size of the criterion variable. To test hypothesis 2 and hypothesis 3, hierarchical multiple regressions were conducted. For hypothesis 2, monitoring, providing consequences, and a multiplicative combination of the two were used as predictors, and employee engagement scores as criteria. For hypothesis 3, monitoring, providing consequences, and providing antecedents were used as predictors, and employee engagement scores as criteria. Additional exploratory zero-order Pearson correlations were conducted between each individual category of supervisory behavior and each of the 12 individual Gallup questions.

Results

Interrater Percentage Agreement

Interrater agreement was assessed by directly comparing scores between a primary rater and a secondary rater for each individual item for every participant. Agreement was scored for 100% of the participants' responses. Comparisons for agreement were made for each participant on an item-by-item basis. Interrater agreement was calculated as a percentage score. Percentages were scored for

scoring units in each item. When the primary rater coded a unit for a respondent on an item (e.g., monitoring), an agreement was noted when the secondary rater coded the same unit. The counts of the units the primary rater coded and the agreements were summed across participants for each of the 16 items. The formula used to calculate percentage agreement for units for an item was number of agreements for unit divided by the number of occurrences of the unit the primary rater coded and then multiplied by 100. Units consisted of the categories of monitoring, providing consequences, and providing antecedents, and the major subcategories for each of them. The units scored on each of the 16 n-basket items differed. A scoring option developed specifically for this scoring system was one referred to as “not identified” (or “ni”). When a participant’s response did not fit any of the definitions or examples given, but the rater judged the response to be a monitor, the rater had the option of coding it as a monitor “not identified” on the code sheet. Responses coded as “not identified” were also considered as a particular unit.

Table 4 shows the interrater agreement percentage scores for each of the categories of behavior, averaged across all items scored. Scores for the major categories were sufficient. For monitoring, the average agreement was 97%, 96.5% for providing consequences, and 82.2% for providing antecedents. These average percentage scores were within the limits of acceptability for a new measure, 80% (Miller, 1997). These results demonstrate that the in-basket resulted in reliable

scoring. All disagreements were reconciled by all three scorers until final agreement was determined. The final reconciled scores were used for the analysis.

Descriptive Statistics for In-basket Categories and Subcategories

Table 5 shows the summary of scores on the in-basket for the categories of monitoring, providing consequences, and providing antecedents, selected subcategories of each, and overall additive and multiplicative combinations. Each category and subcategory is presented with two lines of data in the table. The top line represents scores calculated by dividing the raw score points by the total points possible and then multiplied by 10 for the category or subcategory. The raw scores, or numerator from the scores in the line above, are shown in the bottom line in parentheses. The calculated scores could range from zero to ten, with higher values indicating behaviors that occurred more often than lower values.

As shown in Table 5, the behaviors that were exhibited the most in regard to what was possible were monitoring in total ($M = 4.06$), monitoring via self-report ($M = 3.52$), providing positive when warranted consequences ($M = 3.74$), and providing negative when warranted consequences ($M = 3.17$). The behaviors exhibited least often were providing negative when not warranted consequences ($M = .16$), informing someone out of the contact loop ($M = .35$), providing antecedents overall ($M = .59$), tacking on traditional antecedents ($M = .61$), and relying on antecedents ($M = .51$).

For monitoring, scores ranged from 0.95 to 6.49 (raw point equivalents of 7 and 48, respectively), and the average was 4.06 (30.02 points). For providing consequences, scores ranged from 1.12 (13 raw points) to 4.40 (59 points), and the average was 2.55 (33.72 points). This overall score was comprised of the sum of raw points for positive and negative consequences, divided by the maximum possible points for all consequences (134 points). For providing antecedents overall, total scores ranged from -0.32 to 1.94 (-2 to 12 points), with a mean of 0.59 (3.68 points).

The scores for consequences that were positive, which consisted of points for those that were warranted plus the points for those that were not, resulted in a mean of 3.53 (21.55 raw points), and ranged from 0.49 to 6.56 (3 to 40 points). Scores for consequences that were negative overall (which was equal to the sum of the points for those that were warranted, questionably warranted, and not warranted), ranged from 0 to 3.29 (0 to 22 points), and the mean was 1.74 (12.17 points).

An additional multiplicative score was calculated using the categories of monitoring and providing consequences. For the multiplicative score (calculated by taking the z score conversions of the raw scores for monitoring and providing consequences, adding five, and then multiplying the results), the mean was 10.37 and scores ranged from 2.36 to 21.42.

In-basket Category and Subcategory Relationships

To assess how the major categories and subcategories of each behavior were interrelated, Pearson correlations were calculated between them. Results are depicted in Table 6. None of the major categories were found to be significantly related to each other. Monitoring overall was significantly related to only one subcategory of providing consequences. Respondents who gathered more information about performance were more likely to relay negative evaluations of performance that were questionably warranted ($r = .33, p < .05$). Monitoring via work sample was significantly related to the subcategory of tacking on traditional antecedents ($r = .29, p < .05$). This result indicates that participants who spent more time gathering information by sampling the work of employees were more likely to tack on traditional antecedents. Finally, leaders who spent more time relaying negative evaluations of performance overall were also more likely to rely on providing antecedents to motivate employees ($r = .35, p < .05$).

The correlations between overall categories and subcategories “within them” give an indication of the proportion of the overall score that is comprised of the particular subcategory. The correlation between monitoring via work sampling and monitoring overall was .82 ($p < .01$), showing that work sampling accounted for a high percentage of variability in the monitoring score. Monitoring via self-report was also positively related to monitoring overall ($r = .56, p < .01$).

Interestingly, no significant relationship was found between monitoring via work sampling and monitoring via self-report ($r = .03, p > .05$).

Providing consequences overall was significantly related to providing positive consequences overall ($r = .77, p < .01$), providing negative consequences overall ($r = .54, p < .01$), providing negative warranted consequences ($r = .32, p < .05$), and providing negative questionably warranted consequences ($r = .44, p < .01$).

Providing antecedents was significantly related to both subcategories of behavior. Tacking on traditional antecedents was positively related to providing antecedents overall ($r = .98, p < .01$). This result suggests that a high percentage of variability in providing antecedents can be accounted for by tacking on traditional antecedents. Relying on antecedents was also positively related to providing antecedents overall ($r = .32, p < .05$).

Other relationships to note from Table 6 include those among subcategories for providing consequences and providing antecedents. As shown, positive consequences were not significantly related to negative consequences ($r = -.13, p > .05$), meaning that respondents likely to communicate negative evaluations of performance were not necessarily likely to communicate positive evaluations, and vice versa.

Demographic Variables Relation to In-basket Category Scores

To assess whether differences in in-basket scores might be observed for respondents with different demographic characteristics, correlational analyses were conducted. Bivariate Pearson correlations were computed between the major in-basket categories (i.e., monitoring, providing consequences, providing antecedents, overall additive score, and overall multiplicative score) and the demographic variables assessed, including age, years of education, years in the current organization, years in the current position, years of supervisory experience, days of supervisory training in the current organization, and motivation to perform well on this assessment. See Table 7 for a summary of the results.

Age. The age of respondents was significantly correlated with their scores for providing consequences ($r = -.34, p < .05$). This result means that the older a person was, the less he/she provided consequences on the in-basket. No other scores were significantly related to age of respondents (r s ranges from $-.27$ to $.23, p > .05$).

Education. The years of education of respondents was significantly related to the overall multiplicative score ($r = .31, p < .05$). In other words, multiplying the scores, which does not “allow” high scores on one behavior to compensate for low scores on the other, was positively related to years of education. No other scores were significantly related to years of education (r s ranged from $-.28$ to $.30, p > .05$).

Tenure in organization. The number of years that respondents reported working in the organization was not significantly correlated to any of the behaviors or overall scores (r s ranged from $-.18$ to $-.04$, $p > .05$).

Tenure in position. The total number of years respondents reported being in their current position was not significantly related to scores for monitoring, providing consequences, providing antecedents, or the overall additive and multiplicative scores (r s ranged from $-.09$ to $.29$, $p > .05$).

Supervisory experience. Total years of supervisory experience was also not significantly related to scores for monitoring, providing consequences, providing antecedents, or the overall additive and multiplicative scores (r s ranged from $-.03$ to $.23$, $p > .05$).

Supervisory training. Similar to other demographic variables, the number of days of supervisory training reported by respondents was not significantly related to any of the behaviors or overall scores (r s ranged from $-.18$ to $.12$, $p > .05$).

Overall, it does not appear that respondents with different demographic characteristics responded any differently on the measure, with two exceptions. There was a significant negative correlation between providing consequences scores on the in-basket and respondents' age in years ($r = -.34$, $p < .05$), indicating that the older a person was in a supervisory role, the less he/she provided

consequences. In addition, a significant positive correlation was found between the overall multiplicative score and years of education ($r = .31, p < .05$).

Descriptive Statistics of the Criterion Variable

The Gallup Q12 survey responses by all associates were included in the employee engagement scores for each respondent. See Table 8 for complete results. Overall employee engagement scores were calculated for each respondent by taking the individual mean score for each of the 12 questions, adding them together, and dividing by 12 to get a grand mean. Each of the 12 questions could receive a score between 1.00 and 5.00. The average grand mean employee engagement score was 4.31 (66th percentile) with a range from 3.68 (24th percentile) to 4.87 (95th percentile).

The number of employees who filled the survey out varied greatly between respondents. The average number of employees who filled out the survey for respondents was 36.57 with a low of just 5 survey responses to a high of 112 survey responses. As a result of the large variation in number of employees who participated for each respondent, total employee engagement survey sample size was included in all statistical analyses.

Demographic Variables Relation to Employee Engagement

Analyses similar to those conducted to assess differences in predictor scores for the various demographic variables were performed for the employee

engagement scores. Bivariate Pearson correlations were computed for the employee engagement scores with age, years of education, years in the organization, years in the current position, years of supervisory experience, and days of supervisory training in the current organization. See Table 9 for a summary of results.

Age. No significant relationship was found between respondents' age and employee engagement scores ($r = -.04, p > .05$).

Education. Employee engagement scores were not significantly correlated with respondents' total years of education, $r = .08, p > .05$.

Tenure in the organization. The years respondents had worked in the organization was not significantly related to employee engagement scores ($r = .21, p > .05$).

Tenure in current position. No significant correlations were found between tenure in current position and employee engagement scores ($r = .10, p > .05$).

Supervisory experience. The correlation between years of supervisory experience and employee engagement scores was not significant ($r = -.19, p > .05$).

Supervisory training. The correlation between days of supervisory training in the current organization and employee engagement scores was not significant ($r = -.20, p > .05$).

N size. A significant negative correlation was found between the number of associates who completed the employee engagement survey and the overall

employee engagement scores for each respondent ($r = -.42, p < .01$), indicating that respondents who had a higher number of direct reports fill out the survey had lower overall employee engagement scores for their team.

In conclusion, for employee engagement, none of the participants' demographic characteristics significantly explained the variation in employee engagement scores, with the exception of N size. In general, respondents who had more direct reports received lower overall employee engagement scores.

Hypothesis Testing

Hypothesis 1a, 1b and 1c. Hypothesis 1a predicted that overall monitoring would be significantly related to employee engagement. Overall monitoring was not significantly related to employee engagement scores ($r = .01, p > .05$). (Refer to Table 10.) In addition, monitoring via work sample ($r = -.08, p > .05$) and monitoring via self-report ($r = .11, p > .05$) were also not significantly related to employee engagement scores, leaving hypothesis 1a unsupported.

Hypothesis 1b predicted that overall providing consequences would be significantly related to employee engagement. Providing consequences were also not significantly related to employee engagement scores ($r = .06, p > .05$). No significant relationship was found for positive consequences ($r = .01, p > .05$) or negative consequences ($r = .08, p > .05$). Therefore, hypothesis 1b is also not supported.

Hypothesis 1c concerned the relationship between providing antecedents and employee engagement and predicted the two would be significantly related. Results from the study found that providing antecedents had a significant negative relationship with employee engagement scores ($r = -.32, p < .05$), supporting Hypothesis 1c. In addition, the subcategory of tacking on traditional antecedents was significantly related to employee engagement ($r = -.33, p < .05$). To further assess the relationship between providing antecedents and employee engagement while controlling for N Size, a hierarchical multiple regression was conducted (See Table 11).

The hierarchical multiple regression revealed that at Step one, N Size contributed significantly to the regression model ($F(1, 45) = 9.81, p < .01$) and accounted for 17.9% of the variation in Employee Engagement scores. When added to the regression mode, neither overall monitoring nor overall consequence scores were significant predictors of Employee Engagement. Adding the overall antecedent score to the regression model explained an additional 13.3% of variation in Employee Engagement scores and this change in R^2 was significant, $F(4, 42) = 4.83, p < .01$, further supporting Hypothesis 1c. A leaders' employee engagement score (measured as a grand mean) decreased by .005 points for each additional team member who completed the survey. Employee engagement scores decreased by .196 points for each additional point scored in the providing antecedents category.

Hypothesis 2. Multiple regression analysis was used to test hypothesis 2 which concerns the interaction between monitoring and providing consequences and predicts that the interaction of the two would significantly add to the prediction of employee engagement scores, over and above both alone. This hypothesis was tested in the second step of the hierarchical multiple regression analysis. Due to the significant contribution of N size as a predictor in the previous regression model, it was used as a predictor in the remainder of regression analyses.

As shown in Table 12, the hierarchical multiple regression revealed that at Step one, N size, monitoring overall, and providing consequences overall accounted for 18.1% of the variation in employee engagement and contributed significantly to the regression model $F(3, 43) = 3.18, p < .05$. Adding an interactive term as a predictor did not add significantly to the prediction of employee engagement as hypothesized ($R^2 = .18, F(4, 42) = 2.35, p = .07$).

Hypothesis 3. To determine whether providing antecedents improved the prediction of employee engagement, a hierarchical multiple regression analysis was conducted (see Table 13). The first step was identical to the previous hierarchical multiple regression analysis. Overall monitoring and consequences, and N size were the predictors for employee engagement in Step one. The second step was to add the overall score for providing antecedents to the regression equation. The addition of antecedents overall as a predictor did significantly add to the prediction

of employee engagement by 13.3% ($F(4, 42) = 4.83, p < .01$), supporting hypothesis 3.

Discussion

The primary purpose of the current study was to determine if certain leadership behaviors were indicative of leadership effectiveness, which was measured via an employee engagement survey that followers completed. Previous research has suggested that highly effective leaders spend more time collecting information on performance (monitoring) and providing consequences based on that performance information (Komaki, 1998). Alternatively, then, they spend less time providing instructions, reminders, and training in an effort to motivate their employees. The study functions similar to a descriptive assessment commonly used in behavioral research. We were attempting to answer the difficult question of what specifically do high performing, successful leaders do on a day-to-day basis?

Results from the current study found no significant relationship between leaders' time spent monitoring or providing consequences and employee engagement scores. The study did find that leaders with higher employee engagement scores spent significantly less time providing traditional antecedents like instructions, reminders. An additional unexpected finding was that the size of a leaders' team seemed to play a significant role in leadership effectiveness scores.

The larger a leaders' team, the less likely they were to have high employee engagement scores.

The lack of any significant relationship between employee engagement scores and leaders' time spent collecting performance information and providing consequences was not only contrary to hypothesis 1a and 1b, but much of the previous research on the OMES to date (Brewer, Wilson, & Beck, 1994; Jensen & Komaki, 1993; Komaki, 1986; Komaki, Desselles, & Bowman, 1989; Komaki, Hyttinen, & Immonen, 1991; Komaki, Reynard, & Wallace, 1995). Monitoring occurred often across the assessments in the current study. Average, minimum, and maximum monitoring scores for the current study were in line with Minnich (2007), who used the same assessment and did find significant relationships between monitoring and the studies criterion variable. This suggests it is unlikely a relationship was not detected due to a lack of variation in the behavior itself. Several factors could explain the lack of relationship detected between previously supported behaviors in the OMES and employee engagement.

The OMES and supporting research suggests that monitoring and providing consequences work together. That is, when leaders' monitoring is high, the relationship observed between providing consequences and leadership effectiveness is positive. However, when leaders' monitoring is low, no relationship will be observed between providing consequences and leadership effectiveness. The

accompanying feedback or consequences in general are fundamental to the model's success. The second hypothesis tested for this interaction in the current study, but no significant interaction was detected.

Unlike previous studies, the current study found no significant relationships between participants monitoring and providing consequences for behavior. That is, leaders who spent more time monitoring, did not necessarily spend more time providing consequences and vice versa. For example, one of the more recent studies to use the in-basket found a significant relationship between monitoring overall and providing consequences overall ($r = .32, p < .05$) (Minnich, 2007). However, one interesting relationship between monitoring and a subcategory of negative consequences is worth noting. The current study found a significant positive relationship between monitoring overall and providing questionably warranted negative consequences ($r = .33, p < .05$). This could explain the lack of support for hypothesis 1a and hypothesis 1b. For example, in response to the Director of Art department complaining about a new reimbursement form from the accounting department a leader reached out directly to the leader of accounting and said, "We need to evaluate the new forms that are being used for reimbursement. Let's go back to the old form before we have the discussion so we don't have any more miscommunications with vendors/employees/outside photographers." Because the leader made the assumption the form was actually faulty without verifying or

asking to see it first, this was considered a questionably warranted negative consequence to the leader in accounting. An environment where the majority of feedback is not only negative, but questionably warranted, could result in employees who are less likely to respond favorably to a survey about their company and leadership. Leaders' who tend to provide frequent questionably warranted negative consequences may be discouraged from providing additional monitors based on these findings. An alternative view would suggest that with additional monitors, consequences provided may be less likely to be questionably warranted.

Another possible explanation could lie in the fundamental properties of the criterion variable used in the study. The behavior of monitoring is suggested to be effective in the fourth stage of the motivational process (Komaki, 1998). That is, only when employees already have the resources and training required to do their job and only need motivation to do the job day after day, is monitoring an effective strategy. The employee engagement survey used as the criterion variable may capture feedback from other stages of the motivational process, diluting the chances of a relationship between monitoring and employee engagement being detected. For example, the employee engagement survey used in this study asks employees a question about having the correct resources to do their job. Responses to this question are reflected in the overall employee engagement score for an individual leader. If a leader fails to provide his associates with the required tools needed for

the job, monitoring performance or providing consequences is unlikely to achieve the desired results. In fact, monitoring and providing consequences under these circumstances could have a negative impact on a measure like employee engagement. It might be interesting to identify which questions satisfy the other moderators in the model and then only include leaders who received favorable responses on these questions in the analysis. Correlational analyses were conducted (though not reported) between each of the major behaviors and every individual question on the employee engagement survey. Although no significant relationships emerged, additional regression analyses were not conducted which may have yielded different results.

Another possible explanation for a lack of relationship between monitoring and providing consequences and employee engagement is in a potential confounding variable that was not captured. It is possible that some participants' in the current study may have elements of their individual leadership roles that are incompatible with the behaviors of monitoring and providing consequences. Some leaders may be in positions with higher administrative demands such as paperwork or meetings to attend, reducing the opportunity to spend time monitoring staff performance or providing consequences. It may be that they wish they could spend more time monitoring performance, but competing pressures from their superiors require them to spend their time elsewhere. This type of confound could easily

result in participants' with monitoring and providing consequences scores that are higher than how you might observe them spend their time in their actual leadership role. Future research should consider exploring methods to ensure the real-world opportunity to engage in these behaviors is similar for each participant.

The environment and type of work may have contributed to the lack of significant findings between monitoring, providing consequences, and employee engagement. Daniels and Bailey (2014) distinguish between different types of consequences in the workplace. It may be that certain work environments create more opportunities for employees to come into contact with natural social reinforcers than others. For example, a nurse working in a hospital may be much more likely to come into contact with social reinforcers from the job (i.e., seeing others get healthy, positive feedback from a patient, etc.), than a bank teller helping customers in a bank. If this is true, the frequency of leadership behaviors needed to maintain motivation may be different in the two environments. A leader may need to spend significantly more time monitoring and providing consequences in a bank than in a hospital. Although one of the studies using the OMES was conducted in a human services environment, the majority were conducted in environments where natural social consequences may not be as rich as in a hospital environment. Some of the other settings where the previous research has been conducted include a movie theater, sail boat regatta, and the police force to name a few. The OMES may

only be a useful model for leadership in certain work environments or for certain types of jobs in general. More research is needed to determine what criteria might be useful for determining when to use an in-basket assessment for leadership development or selection purposes.

A final consideration for a lack of relationship between monitoring performance and providing consequences could be that these behaviors are simply not as important for leadership effectiveness as previously hypothesized. It is possible that with further research on the model, monitoring and providing consequences will prove to be less critical to leadership effectiveness than previously thought. Although this is unlikely to be the case due to the previous studies supporting the model and behaviors to date, it is an important possibility to keep in mind. Particularly when discussing studies that did not demonstrate the hypothesized relationships.

The current study found that leaders' who spent less time providing traditional antecedents were more likely to have higher employee engagement scores than leaders' who provided more traditional antecedents to employees. This significant relationship between the 'antecedent' behavior category and employee engagement is the first of its kind. No previous studies utilizing the in-basket have demonstrated a significant relationship with this category on the in-basket assessment and any leadership effectiveness criteria.

Table 14 shows the difference in responses for top and bottom rated leaders in terms of employee engagement scores. The item asks the participant to respond to an email in which the Director of Human Resources is complaining about another department not showing up to a scheduled training, resulting in the trainer to leave. The email includes the map and directions provided to the department. When looking at the map and instructions, it is clear the map has errors leading to the wrong building and the written instructions are not correct either. A leader with a high employee engagement score responded by expressing confusion with the map and directions and pointed out the errors in both. A leader with a low employee engagement score responded with what qualified as a tacked on traditional antecedent. The response suggested the Director of Human Resources add additional and seemingly unnecessary information to the map and did not mention the obvious errors in the map and/or instructions. He/she wrote, "I would suggest adding street numbers and an address to the training center on this map, so that the trainees could use GPS/map program to help them locate the facility." Not only did the leader fail to recognize any errors in the map, they persisted with incorrect information as a means to solve the problem. Although there are clear differences at first glance, the behaviors measured did not capture these differences in this case.

The third and final hypothesis predicted that the addition of leaders' antecedent scores would aid in the prediction of leadership effectiveness on top of monitoring and providing consequences alone. Although hypothesis 3 was supported in the current study, the results need to be interpreted with caution. Even though the addition of providing antecedents significantly added to the prediction of leadership effectiveness, both monitoring and providing consequences remained as insignificant variables in the model.

The significant relationship discovered between team size and employee engagement scores has several implications. The finding suggests that highly effective leaders have smaller teams, in general. As a leaders' team grows, it may become more difficult to be effective or successful. Previous studies have suggested that managers should have no more than seven or eight direct reports, though actual numbers vary by industry (Davison, 2003). It may be that as a leaders' team grows in size, monitoring and providing consequences becomes a more challenging strategy to execute. The leaders' in the current study had an average of 36.57 employees who filled out a survey and some had up to 112 employees. It is unlikely that a leader could adequately provide monitors and consequences effectively for 36 employees and especially unlikely for 112 employees. If this is the case, then an additional moderator could be added to the OMES regarding number of team members. Although it seems only logical that number of employees a leader is

directly responsible for might impact leadership effectiveness, further research is needed to determine if this is due to not being able to monitor and provide consequences as often or some other factor.

In sum, monitoring and providing consequences were not found to be significantly related to leadership effectiveness as hypothesized. However, providing antecedents was found to be significantly related to leadership effectiveness. In addition, number of employees who rated their supervisor was also significantly negatively related to leadership effectiveness.

Limitations

Several limitations in the current study are worth mentioning. The first limitation is regarding the criterion variable selected. As mentioned previously, several questions on the employee engagement survey that was used to measure leadership effectiveness may have exceeded the scope of the model and leadership behaviors under study. The OMES is hypothesized to be impacted by five moderator variables mentioned earlier. These moderators include characteristics of the follower, characteristics of the leader, resources of the organization, stage of motivational process, and type of task (Komaki, 1998). No measures were available to know whether each of these moderators was satisfied in the current study. It could be that certain departments were short on resources, limiting the impact that monitoring and providing consequences might have.

In addition, several of the questions on the employee engagement survey would likely be more suitable measures for one or more moderator variables. For example, one question on the survey asked employees if they had the necessary resources to do their job. This may have resulted in higher or lower than appropriate leadership effectiveness scores given the behaviors under study. The current study used the employee engagement scores as the measure for leadership effectiveness largely due to it being one of the only consistent measures associated with each leader. Selecting specific questions from the survey to use alone as the measure of leadership effectiveness may have yielded different results.

A second limitation of the current study was the limited information regarding the leadership structure within the organization's departments. Although the employee engagement surveys are linked directly to the acting manager of the person filling it out, many departments within the hospital system have assistant managers or supervisors who assist in managing the department but are not linked to the employee engagement scores directly. Not knowing which participants in the study have additional leadership support threatens the internal validity of the overall findings.

Another limitation is with the use of an employee engagement metric as a criterion for leadership effectiveness. As mentioned earlier, a sizeable amount of data has been published to suggest that better employee engagement scores lead to

better organizational outcomes (Harter, Schmidt, Agrawal, Plowman, & Blue, 2016). Although these findings are impressive, exceptions likely exist within the data. No evidence was provided from the organization under study that suggests departments with higher employee engagement scores tended to have better performance outcomes.

The amount of participants in the current study is also a major limitation. It is likely that the statistical power of the results in this study are low. A power analysis conducted prior to the study suggested a minimum of 52 participants for a moderate power level. The current study was five participants short. With a low power, the likelihood of a Type II error is raised. It is possible that an effect in the study existed but went undetected.

One limitation in the current study pertains to the in-basket assessment. The in-basket offers a more practical method of assessing leadership behavior in a workplace. However, a difficulty in scaling or valuation of behaviors is a limitation. During a live observation of a leaders' performance, the measurement of a given behavior is a simple percentage of time. If a leader delivered 2 consequences in 10 intervals of time, it would count the same as if they delivered 2 monitors in 10 intervals of time. The in-basket is different in that, point values were assigned to leadership behaviors based on the quality of the behavior rather than its occurrence over time. A point value of 1 through 5 was assigned to a positive consequence,

whereas a score from -6 to +6 was assigned to a negative consequence. It is unclear if the behaviors in the current scoring system are valued appropriately. It could be that subtracting six points for certain negative consequences is too high. Further consideration should be given to scoring procedures used with the in-basket assessment.

In addition to a difficulty in scoring and scaling of behaviors, an imbalance was observed between providing antecedents and providing monitors and consequences. That is, there are many more opportunities to provide monitors and consequences than provide antecedents. Future researchers should consider standardizing the scores for each category to see if different patterns emerge in the analyses.

An additional limitation to the in-basket is that the measure of leadership behavior relies on self-report. One of the reasons the field of behavior analysis utilizes direct observation is to limit the bias and subjectivity involved in self-report measures. With the limited studies conducted which validate the in-basket assessment, it is possible that responses are biased in nature and not a true reflection of how any leader may actually spend their time.

Future Research

Several avenues exist for future researchers to explore. Additional studies conducted in a hospital setting with the in-basket are warranted. The current study

was unable to replicate significant relationships between monitoring and providing consequences and employee engagement. Future researchers should conduct a study similar to the current, but use one of the measures of motivational effectiveness that previous research on the OMES has used. This might help determine if there are fundamental differences in what makes for an effective leader in a hospital and other high demand environments.

The use of an employee engagement survey as a measure of leadership effectiveness could be improved. As mentioned earlier, some questions might not have been appropriate for the model and behaviors under study. Future researchers should consider creating an “index” of questions from an employee engagement survey that might better represent leadership effectiveness in this context. For example, a question asking employees if they have received recognition recently might be impacted more by the behaviors in the OMES than a question that asks about having the appropriate resources to do the job. Another possible leadership criterion could be an index that includes a measure of employee satisfaction and customer satisfaction. These two scores would likely have separate measures but be combined to provide each leader with an overall “satisfaction index” which could serve as a more dynamic measure of leadership effectiveness.

More research needs to be conducted to demonstrate construct validity of the in-basket. Future researchers should conduct additional studies similar to the

original study involving the OSIBA (Komaki et al., 1990), to provide further evidence that the behaviors thought to be measured on the OSIBA are actually being measured. Komaki et al. (1990) did find a significant relationship between monitoring and providing consequences as measured on the in-basket with observations of performance on the job. This relationship has yet to be explored with providing antecedents as it is a relatively newer behavior to be measured from the in-basket. Future researchers could look to validate that providing antecedents on the in-basket captures an accurate representation of when this behavior may occur in the actual environment the leader works in.

Future researchers should consider further evaluating the leadership behaviors which result in higher levels of leadership effectiveness with a similar criterion. The current study effectively demonstrated what a leader should “not do” to be effective. Further exploratory analyses could be conducted under similar circumstances to determine if any patterns emerge in the behavior of leaders with higher leadership effectiveness scores. Perhaps a behavioral construct not being considered such as ensuring employees have adequate resources or removing barriers could be captured in an in-basket assessment, which is not currently being measured.

Additional studies should be conducted to further analyze the impact of management scope on leadership performance. The significant negative

relationship between the number of employees who filled out the survey and the measure utilized for leadership effectiveness was interesting and warrants further exploration. Although no significant relationships were found between the behaviors measured on the in-basket and team size, future research could explore the impact team size has on a leader's ability to perform behaviors in the model. A 'break point' could exist in a manager's team size in which she is no longer capable of performing consistent quality monitoring and consequences. Future researchers should use the in-basket to help determine where this 'break point' might be. Future research should also consider other factors that might explain why a large team could result in lower employee engagement scores.

Future researchers should consider taking an experimental approach to the study of the Operant Model. One possibility is to use a coaching methodology to change their monitoring and providing consequences behavior over time and observe the impact this has on a specified criterion variable. The OSTI or OSIBA could be used to measure the change in leadership behavior at an individual level. Another possibility would be to conduct a between-groups study. One group could receive training which encourages frequent monitoring of performance and how to provide performance consequences. The other group could serve as a control group. The OSIBA or OSTI could be used to determine if differences in leadership behavior existed following the training.

In summary, the current study has shown that a specific supervisory behavior grounded in theory and measured via the in-basket and leaders' team size were negatively related to employee engagement scores. Although these findings do add to the leadership literature, much more leadership research is needed in hospital settings and the field of OBM overall.

The science of behavior analysis utilizes a scientific approach to the study of behavior in any environment. This approach to the study of behavior provides a unique opportunity for the field to help advance our understanding of leadership behavior across the board. Unfortunately, few empirical studies have been conducted which have attempted to further advance our understanding of what effective leaders do differently than ineffective leaders. Researchers in the field of behavior analysis should consider pursuing any opportunity to study leadership behavior that becomes available.

With more research, the in-basket assessment could help further define what "effective leadership" looks like. A practical tool that captures leadership behavior could provide organizations a better means to evaluate and develop their leaders' skills over time. Additionally, the in-basket could have further applications in the selection of an individual for a leadership position. With better tools to evaluate if a candidate possesses the necessary skills to be successful in a leadership role, many wasteful situations for both the organization and leader can be avoided entirely.

References

- Bakker, A. B., Albrecht, S. L., & Leiter, M. P. (2011). Key questions regarding work engagement. *European Journal of Work and Organizational Psychology, 20*(1), 4-28.
- Bakker, A. B. & Bal, P. M. (2010). Weekly work engagement and performance: A study among starting teachers. *Journal of Occupational and Organizational Psychology, 83*(1), 189-206.
- Bakker, A. B., Hakanen, J. J., Demerouti, E., & Xanthopoulou, D. (2007). Job resources boost work engagement, particularly when job demands are high. *Journal of Educational Psychology, 99*(2), 274-284.
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.
- Bass, B. M. (2008). *Handbook of leadership: Theory, research, and managerial applications* (4th ed.). New York: Free Press.
- Beck, R. & Harter, J. (2015). Managers account for 70% of variance in employee engagement. *Gallup Business Journal*.
- Brass, D. J. & Oldham, G. R. (1976). Validating an in-basket test using an alternative set of leadership scoring dimensions. *Journal of Applied Psychology, 61*(5), 652-657.

- Brewer, N., Wilson, C., & Beck, K. (1994). Supervisory behavior and team performance amongst police patrol sergeants. *Journal of Occupational and Organizational Psychology*, 67(1), 69-78.
- Christian, M. S., Garza, A. S., & Slaughter, J. E. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology*, 64(1), 89–136.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2014). Applied behavior analysis (Second ed.). Harlow: Pearson Education.
- Daniels, A. C. (1994). *Bringing out the best in people*. New York: McGraw-Hill.
- Daniels, A. C. (2009). *Oops! 13 management practices that waste time and money (and what to do instead)*. Atlanta, GA: Performance Management.
- Daniels, A. C., & Bailey, J. S. (2014). Performance management: changing behavior that drives organizational effectiveness (Fifth ed.). Atlanta, GA: Performance Management Publications.
- Daniels, A. C., & Daniels, J. E. (2007). *Measure of a leader*. New York: McGraw-Hill.
- Davison, B. (2003). Management span of control: How wide is too wide? *Journal of Business Strategy*, 24(4), 22-29.
- Gallup Employee Engagement Center. (n.d.). Retrieved March 29, 2017, from <https://q12.gallup.com/Help/en-us/About>

- Gallup, I. (2017, February 15). Gallup Daily: U.S. Employee Engagement. Retrieved March 29, 2017, from <http://www.gallup.com/poll/180404/gallup-daily-employee-engagement.aspx>
- Gill, R. W. T. (1979). The in-tray (in-basket) exercise as a measure of management potential. *Journal of Occupational Psychology*, 52(3), 185-197.
- Hackman, J. R. & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology*, 60(2), 159-170.
- Harter, J. K. Schmidt, F. L., Agrawal, S., Plowman, S. K. & Blue, A. (2016). The relationship between engagement at work and organizational outcomes. *Gallup Business Journal*.
- Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology*, 87(2), 268-279.
- Houmanfar, R. A., Alavosius, M. P., Morford, Z. H., Herbst, S. A., & Reimer, D. (2015). Functions of organizational leaders in cultural change: financial and social well-being. *Journal of Organizational Behavior Management*, 35(1-2), 4-27.

- House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt and L. L. Larson (Eds.), *Leadership: The cutting edge* (pp. 189-207). Carbondale Southern Illinois University Press.
- Jensen, M. & Komaki, J. L. (1993, May). Beware of too many directives: Spotting a suppressor variable in the operant model of effective supervision. In J. Cannon-Bowers (Chair), *Optimizing team performance through team leader behavior*. Symposium conducted at the conference of the Society of Industrial and Organizational Psychology, San Francisco, CA.
- Johnson, G. A. (1975). The relative efficacy of stimulus vs. reinforcement control for obtaining stable performance change. *Journal of Organizational and Human Performance*, 14(3), 321-341.
- Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (Third ed.). New York: Routledge.
- Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, 33(4), 692-724.
- Kaiser, R. B., Hogan, R. & Craig, S. B. (2008). Leadership and the fate of organizations. *American Psychologist*, 63(2), 96-110.
- Komaki, J. L. (1986). Toward effective supervision: An operant analysis and comparison of managers at work. *Journal of Applied Psychology*, 71(2), 270-278.

- Komaki, J. L. (1998). *Leadership from an operant perspective*. London, England: Routledge.
- Komaki, J. L. & Citera, M. (1990). Beyond effective supervision: identifying key interactions between superior and subordinate. *Leadership Quarterly*, 1(2), 91-105.
- Komaki, J. L., Coombs, T., & Schepman, S. (1991). Motivational implications of reinforcement theory. In R. M. Steers and L. W. Porter (eds.), *Motivation and work behavior* (5th edn, pp. 87-105). New York: McGraw-Hill.
- Komaki, J. L., Desselles, M. L., & Bowman, E. (1989). Definitely not a breeze: Extending an operant model of effective supervision to teams. *Journal of Applied Psychology*, 74(3), 522-529.
- Komaki, J. L., Hyttinen, M., & Immonen, S. (1991, August). Cross-cultural research in work and leader behavior. In K. R. Thompson (Chair), *Managing across cultures from a behavioral perspective: Theory and research*. Symposium presented at the meeting of the Academy of Management, Atlanta.

- Komaki, J. L., Newlin, M. H., & Desselles, M. L. (1990, April). Walking on the wild side: Criterion related validation of an in-basket exercise of supervisory behaviors. In M. H. Newlin (Chair), *Simulated performance assessment: Fact or Fantasy?*. Symposium presented at the conference of the society for Industrial and Organizational Psychology, Miami, FL.
- Komaki, J. L., Minnich, M. L. R., Grotto, A. R., Weinshank, B., & Kern, M. J. (2011). Promoting critical operant-based leadership while decreasing ubiquitous directives and exhortations. *Journal of Organizational Behavior Management*, 31(4), 236-261.
- Komaki, J. L., Reynard Minnich, M., Lee, C., & Wallace, J. (1997, May). Apples and oranges at sea: Skippers' leadership styles make a difference. In B. P. Niehoff (Chair), *Research on performance monitoring: Refinements and boundaries*. Symposium presented at the conference of the Society for Industrial and Organizational Psychology, St. Louis.
- Komaki, J. L., Zlotnick, S., & Jensen, M. (1986). Development of an operant-based taxonomy and observational index of supervisory behavior. *Journal of Applied Psychology*, 71(2), 260-269.
- Krapfl, J. E. & Kruja, B. (2015). Leadership and culture. *Journal of Organizational Behavior Management*, 35(1-2), 28-43.

- Locke, E. A. & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Ludwig, T. D. & Frazier, C. B. (2012). Employee engagement and organizational behavior management. *Journal of Organizational Behavior Management*, 32(1), 75-82.
- Macey, W. H., Schneider, B., Barbera, K. M., & Young, S. A. (2009). *Employee engagement: Tools for analysis, practice, and competitive advantage*. Malden, MA: Wiley-Blackwell.
- Maslach, C., & Leiter, M. P. (2008). Early predictors of job burnout and engagement. *Journal of Applied Psychology*, 93(3), 498–512.
- Mauno, S., Kinnunen, U. & Ruokolainen, M. (2007). Job demands and resources as antecedents of work engagement: A longitudinal study. *Journal of Vocational Behavior*, 70(1), 149-171.
- Methot, L. L., Williams, W. L., Cummings, A., & Bradshaw, B. (1996). Measuring the effects of a manager-supervisor training program through the generalized performance of managers, supervisors, front-line staff, and clients in a human service setting. *Journal of Organizational Behavior Management*, 16(2), 3-34.
- Miller, L. K. (1997). *Principles of everyday behavior analysis* (3rd ed.). Belmont, CA: Wadsworth.

- Moon, K. & Oah, S. (2013). A comparison of the effects of feedback and prompts on safe sitting posture: Utilizing an automated observation and feedback system. *Journal of Organizational Behavior Management*, 33(2), 152-162.
- Murray, D. (1995). *A democracy of despots*. Montréal: McGill-Queen's University Press.
- Nitsch, J. R. (1976). *Zur Theorie der sportlichen Beanspruchung [The theory of strain in sport]*. In J. R. Nitsch and I. Udris (Eds.), *Beanspruchung im Sport* (pp. 15-41). Bad Homburg: Limpert.
- Rich, B. L., LePine, J. A., & Crawford, E. R. (2010). Job engagement: Antecedents and effects of job performance. *Academy of Management Journal*, 53(3), 617-635.
- Reynard, M. & Komaki, J. L. (1995, May). *The music of management: Rhythms of effective leaders*. Poster presented at the conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Reynard Minnich, M. L. (2007). The role of performance directives in predicting supervisory effectiveness. Dissertation.
- Robison, J. (2006). In praise of praising your employees; Frequent recognition is a surefire – and affordable – way to boost employee engagement. *Gallup Management Journal*.

- Saks, A. M. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology*, 21(7), 600-619.
- Saks, A. M. & Gruman, J. A. (2014). What do we really know about employee engagement? *Human Resources Development Quarterly*, 25(2), 155-182.
- Schaufeli, W. B. & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, 25(3), 293-315.
- Schaufeli, W. B., Bakker, A. B., & Van Rhenen, W. (2009). How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism. *Journal of Organizational Behavior*, 30(7), 893-917.
- Schaufeli, W. B., Salanova, M., Gonzalez-Roma, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3(1), 71-92.
- Schippmann, J. S., Prien, E. P., & Katz, J. A. (1990). Reliability and validity of in-basket performance measures. *Personnel Psychology*, 43(4), 837-859.
- Schneider, S. (2012). *The science of consequences: how they affect genes, change the brain, and impact our world*. Amherst, NY: Prometheus Books.
- Scott, W. E., Jr. & Podsakoff, P. M. (1985). *Behavioral principles in the practice of management*. New York: Wiley.

- Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press;
London.
- Skinner, B. F. (1974). *About behaviorism*. New York: Vintage.
- Tucker, E. & Lam, S. (2014). Dynamic leadership – a leadership shortage solution.
Strategic HR Review, 13(4-5), 199-204.
- Vroom, V. H. & Jago, A. G. (2007). The role of the situation in leadership.
American Psychologist, 62(1), 17-24.
- Xanthopoulou, D., Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2009a).
Reciprocal relationships between job resources, personal resources, and
work engagement. *Journal of Vocational Behavior*, 74(3), 235-244.
- Yukl, G. (2013). *Leadership in organizations* (Eighth ed.). Harlow (England):
Pearson Education.

Table 1. *Categories and Subcategories of Behavior in the OSIBA: Definitions and Examples*

CATEGORY/ Subcategory	Definition	Examples		
		OSIBA item 1 Complaint from employee in Dept. A about difficulty with a new form devised by Dept. B	OSIBA item 2 Relay of client criticism about service in Dept. C	OSIBA item 3 Concern expressed by employee of Dept. D about the shoddy work of Dept. E
MONITORING	Gathering information about performance			
Work sample	Indicating interest in observing workers in action and/or examining work products	<i>Please forward a copy of both the old and new forms to me, I will review.</i>	<i>I would like you to personally visit the department... [and] to call.... See how you are treated and then report back to me and we will discuss it further.</i>	<i>I will be meeting with [Dept. E Mgr.]....Send me the sample of [work] in question.</i>
Self-report	Asking person for more information about performance	<i>Please let me know about what changes have been made in the new...forms— were the changes necessary and were</i>	<i>[To Dept. C Dir.][A client] is very angry....Note: While I am aware of the complaint, is there any reason to</i>	<i>[To Dept. E Dir.]Is there a problem with [work in your department]? Please confer with [Dept. E Mgr.] [about it].</i>

<i>instructions sent out with the new forms?</i>					<i>believe it is true? Who is handling their account?</i>
TABLE 1 (Continued)					
CATEGORY/ Subcategory	Definition	OSIBA item 2 Relay of client criticism about service in Dept. C	OSIBA item 4 Material sent with mistake	OSIBA item 5 Discussion by Dept. F's Dir. About Dept. G's lack of follow through, including misleading attachment	
PROVIDING CONSEQUENCES	Communicating knowledge of performance				
Positive or	Recognizing accomplishment or effort	<i>Thanks for bringing me up to speed on this. I will call [client] myself to assure them that their concerns are important to us...</i>	<i>The [material] look[s] good. The [error] I'm not so sure. Stop the presses!</i>	<i>I appreciate the efforts you made in providing... trainees with [attachment]. However there is a bit of confusion when reviewing [it].</i>	

Neutral	Expressing neither approval nor disapproval of performance	<i>...I've addressed your concerns with [Dept. C] and will do my best to correct the problem...</i>	----	<i>[To Dept. G Dir.] I heard there was a problem....Please let me know if I can help to solve this problem. I asked [Dept. F Dir.] to [provide additional assistance].</i>
Negative	Pointing out error seen first-hand, but in a way that is not sarcastic, abrupt, or demeaning	----	<i>Hold up! You have a [mistake]. [described] Please ...review.... These can be costly mistakes and I am sure you can see, unprofessional.</i>	<i>Pardon me if I point out to you the [error in your attachment]. I can see where [Dept. G] could have had a problem.</i>

TABLE 1 (Continued)

CATEGORY/ Subcategory	Definition	OSIBA item 2 Relay of client criticism about service in Dept. C	OSIBA item 4 Material sent with mistake	OSIBA item 6 Employee provides status report on preparations
PROVIDING ANTECEDENTS	Conveying expectations of performance via directives, instructions, reminders, exhortations, or persuasions			
Tacking on traditional antecedents	Giving directives, reminders, instructions when neither necessary nor requested	<i>[Dept. C Dir.] will need to handle him directly. [She] should call immediately and give them her pager and cell numbers.</i>	<i>As a suggestion, how about adding color and something to call attention to [it]?</i>	<i>OK [sender], let's keep it up, like you said, "Two short months away!" [To sender's Dept. Dir.] [memo sender] updated me on the [project]. Time is running short, maybe you need to get [employee name] and [employee name 2] some help. We need these tasks finalized by the end of this week please. Update me Friday morning.</i>

Using antecedents for motivation	Relying on(or recommending) directives or exhortations alone (without monitoring or providing consequences) in an attempt to promote performance	<i>[To Dept. C Dir] We have received a complaint....The client claims [problems with] the service. Plz review the company policies with your employees if they need more training....Plz contact [HR] to set up training times. We need to keep this client happy. Thanks!!!"</i>	<i>[To sender's Dept. Dir.] [memo sender] is sending her [material] along. Please have your staff make sure everything is spelled correctly before we print. Check twice, print once makes it a better job.</i>	<i>[To Dept. E Dir.]Is there a problem with [work in your department]? Please confer with [Dept. E Mgr.] [about it].</i>
-------------------------------------	--	---	---	--

Note. Reprinted from "Promoting critical operant-based leadership while decreasing ubiquitous directives and exhortations", Komaki, J. L., 2011, *Journal of Organizational Behavior Management*, 31(4), pp.

Table 2. *Summary of Demographic Information*

Demographic variable	<i>N</i>	Minimum	Maximum	<i>M</i>	<i>SD</i>
Age	46	29	62	48.1	8.5
Education	43	12	28	16.3	2.3
Tenure in organization	47	1	30	11.1	8.1
Tenure in position	47	0	14	4	3
Supervisory experience	46	2	30	10.1	7.1
Supervisory training ^a	30	0	60	13.2	13.7

Note. Except where indicated, the unit of measurement is years.

^a Unite of measurement is days.

Table 3. *Points Possible for Categories and Subcategories of Behavior*

Item Number	Monitoring		Providing Consequences					Providing Antecedents	
	Work Sample	Self-report	Positive	Not Warranted	Negative Warranted	Questionably Warranted	Not Warranted	Tacking on Unessential Ones	Relying on Them
1	12	4	4	1	6		1	5	
2		4	4			4			2
3	6	2	4			3		5	
4	6	2	4			3		5	
5	6	2	3	1	6	3		5	
6		4	5		6	3		5	
7	6		2	1	6	3		5	
8			3	1	6	3		5	
9		4	5			4			2
10		6	4					5	
11			2			4			2
12		4	3					5	
13	6	2	4					5	
14			2					2	
15	6	2	4			4			2
16		4	4			7	1		2
Total	48	40	57	4	30	41	2	52	10

Note. Points possible are derived from the score sheets for each item. A respondent could receive additional points for behaviors that were not designated on the score sheet.

Table 4. *Average Interrater Agreement Scores for Major Categories*

Category	Interrater Agreement	
	Frequency ^a	Percentage
Monitors	428/443	96.6%
Consequences	583/604	96.5%
Antecedents	60/73	82.2%

Note: Agreement refers only to occurrence of categories. ^a Frequency = number of agreements divided by number of occurrences.

Table 5. *Summary of In-basket Scores*

Category/ Subcategory 1/ Subcategory 2/ Subcategory 3	Mean	SD	Minimum	Maximum
Monitoring	4.06 (30.02)	1.24 (9.17)	0.95 (7)	6.49 (48)
<i>Work sample</i>	2.89 (13.89)	1.65 (7.92)	0.00 (0)	6.25 (30)
<i>Self-report</i>	3.52 (14.09)	1.22 (4.88)	0.50 (2)	6.00 (24)
Providing consequences	2.55 (33.72)	0.72 (9.60)	1.12 (13)	4.40 (59)
<i>Positive</i>	3.53 (21.55)	1.34 (8.20)	0.49 (3)	6.56 (40)
When warranted	3.74 (21.30)	1.42 (8.10)	0.53 (3)	6.84 (39)
Simple	0.57 (0.85)	0.61 (0.91)	0.00 (0)	2.00 (3)
Letting sender know of action	0.77 (2.09)	0.78 (2.11)	0.00 (0)	3.33 (9)
Relaying problem non-judgmentally	1.23 (1.72)	1.60 (2.23)	0.00 (0)	6.43 (9)
Acknowledging bearer of bad news	1.51 (4.09)	1.38 (3.72)	0.00 (0)	6.67 (18)
Informing someone out of contact loop	0.35 (1.11)	0.67 (2.16)	0.00 (0)	2.50 (8)
Broadcasting to multiple parties	0.87 (1.21)	1.68 (2.35)	0.00 (0)	7.14 (10)
When not warranted	0.73 (0.26)	1.39 (0.49)	0.00 (0)	5.71 (2)
<i>Negative</i>	1.74 (12.17)	0.85 (5.99)	0.00 (0)	3.29 (22)
When warranted	3.17 (9.51)	1.73 (5.18)	0.00 (0)	6.33 (19)
When questionably warranted	0.77 (2.64)	0.91 (3.65)	-0.49 (-2)	3.17 (11)
When not warranted	0.16	0.77	0.00	4.17

	(0.02)	(0.15)	(0)	(1)
Providing antecedents	0.59	0.60	-0.32	1.94
	(3.68)	(3.71)	(-2)	(12)
<i>Tacking on traditional antecedents</i>	0.61	0.68	-0.38	2.31
	(3.17)	(3.54)	(-2)	(12)
<i>Relying on antecedents</i>	0.51	0.78	0.00	2.00
	(0.51)	(0.78)	(0)	(2)
Overall multiplicative (monitoring X consequences)	10.37	4.64	2.36	21.42

Note. N = 47. Numbers shown represent raw scores divided by points possible, multiplied by 10. (Higher values indicate that behaviors occurred more often).

Table 6. *Intercorrelations Between Major Categories and Subcategories of Behavior in the OSIBA*

Category/ Subcategory 1/ Subcategory 2/ Subcategory 3	1	1a	1b	2	2a	2a1	2a2	2b	2b1	2b2	2b3	3	3a	3b
1. Monitoring	1													
1a. Work sample	.82**	1												
1b. Self-report	.56**	.03	1											
2. Consequences	.17	.10	.13	1										
2a. Positive	.15	.07	.11	.77**	1									
2a1. Warranted	.17	.08	.12	.79**	.998**	1								
2a2. Not warranted	-.18	-.15	-.09	-.13	.23	.17	1							
2b. Negative	.06	.06	.06	.54**	-.13	-.10	-.50**	1						
2b1. Warranted	-.16	-.03	-.19	.32*	-.22	-.20	-.39**	.79**	1					
2b2. Questionably warranted	.33*	.15	.37**	.44**	.10	.12	-.30*	.55**	-.07	1				
2b3. Not warranted	-.16	-.12	-.11	.06	-.03	-.04	.13	.14	-.02	.23	1			
3. Antecedents	.13	.28	-.18	-.03	-.03	-.03	.00	-.01	-.07	.08	-.06	1		
3a. Tacking on traditional	.16	.29*	-.14	-.05	.00	.00	.06	-.08	-.12	.03	-.08	.98**	1	
3b. Relying on antecedents	-.13	.01	-.22	.10	-.15	-.13	-.29*	.35*	.26	.22	.10	.32*	.11	1

** $p \leq .01$. * $p \leq .05$.

Table 7. *Correlations Between Selected Demographic Variables and In-basket Scores*

Demographic Variable	Monitoring	Providing Consequences	Providing Antecedents	Overall Additive	Overall Multiplicative
Age	-.11	-.34*	.23	-.24	-.27
Education	.24	.23	-.28	.30	.31*
Tenure in organization	-.04	-.18	-.04	-.12	-.15
Tenure in position	-.07	-.06	.29	-.09	-.11
Supervisory experience	-.03	.15	.23	.04	.04
Supervisory training	-.18	-.09	.12	-.16	-.12

Note. Except where indicated, the unit of measurement is years.

^a Unit of measurement is days.

* $p < .05$, two-tailed.

Table 8. *Descriptive Statistics of Criterion Variable*

	Mean	Standard Deviation	Minimum	Maximum
Gallup Overall Grand Mean	4.31	.32	3.68	4.87
Gallup Overall Percentile Rank	66.47	20.14	24	95
N Size	36.57	27.30	5	112

Table 9. *Correlations Between Selected Demographic Variables and Employee Engagement Scores*

Demographic Variable	Employee Engagement
Age	-.04
Education	.08
Tenure in organization	.21
Tenure in position	.10
Supervisory experience	-.19
Supervisory training ^a	-.20
N Size	-.42**

Note. Except where indicated, the unit of measurement is years.

^a Unit of measurement is days.

** $p < .01$, two-tailed.

Table 10. *Correlations Between In-basket Category and Subcategory Scores and Employee Engagement Scores*

Category/ Subcategory 1/ Subcategory2/ Subcategory3	Overall Employee Engagement Score
Monitoring	.01
<i>Work sample</i>	-.08
<i>Self-report</i>	.11
Providing consequences	.06
<i>Positive</i>	.01
When warranted	-.01
Simple	.13
Letting sender know of action	.10
Relaying problem non-judgmentally	-.09
Acknowledging bearer of bad news	-.06
Informing someone out of contact	-.04
loop	
Broadcasting to multiple parties	-.19
When not warranted	.18
<i>Negative</i>	.08
When warranted	.12
When questionably warranted	-.02
When not warranted	.10
Providing antecedents	-.32*
<i>Tacking on traditional antecedents</i>	-.33*
<i>Relying on antecedents</i>	-.03

* $p < .05$ (2-tailed).

Table 11. *Hierarchical Multiple Regression Results Using Antecedents and N Size as Predictors of Employee Engagement Scores*

Variable	β	R^2	ΔR^2
Step 1		.18	
N Size	-.42**		
Step 2		.32	.14**
N Size	-.46***		
Antecedents			
(Overall)	-.37**		

** $p < .01$, *** $p < .001$

Table 12. *Hierarchical Multiple Regression Results Using N Size, Monitors, Consequences, and an Interactive Term as Predictors of Employee Engagement Scores*

Variable	β	R^2	ΔR^2
Step 1		.18	
N Size	-.43**		
Monitoring			
(Overall)	-.1		
Consequences			
(Overall)	.01		
Step 2		.18	.002
N Size	-.42**		
Monitoring			
(Overall)	-.18		
Consequences			
(Overall)	-.11		
Interactive Term (M			
x C)	.20		

** $p < .01$

Table 13. *Hierarchical Multiple Regression Results Using N Size, Monitors, Consequences, and Antecedents as Predictors of Employee Engagement Scores*

Variable	β	R^2	ΔR^2
Step 1		.18	
N Size	-.43**		
Monitoring			
(Overall)	-.1		
Consequences			
(Overall)	.01		
Step 2		.32	.13**
N Size	-.47***		
Monitoring			
(Overall)	-.01		
Consequences			
(Overall)	-.02		
Antecedents			
(Overall)	.37**		

** $p < .01$, *** $p < .001$

Table 14. *Tacking on Traditional Antecedents in Response to an OSIBA Item from a Leaders with High and Low Employee Engagement Scores*

Item 8	
Item description: Email from director of HR complaining that the marketing department did not show up for a training on time resulting in the trainer leaving before anyone arrived. The map and directions are included in her email, however, there are errors and discrepancies between the map and written instructions.	
Responses	
Leader with High Employee Engagement	Leader with Low Employee Engagement
To Director of HR: So, I'm confused. Was the training at the Michelin Building or the Training Center? Your directions say, Michelin Building but your map tends to direct to the Training Center. I suggest that the training just be rescheduled and new directions and map sent out. Just start over.	To Director of HR: <i>I would suggest adding street numbers and an address to the training center on this map, so that the trainees could use GPS/map program to help them locate the facility.</i>
<i>Note.</i> Tacking on traditional antecedents is shown in italics.	

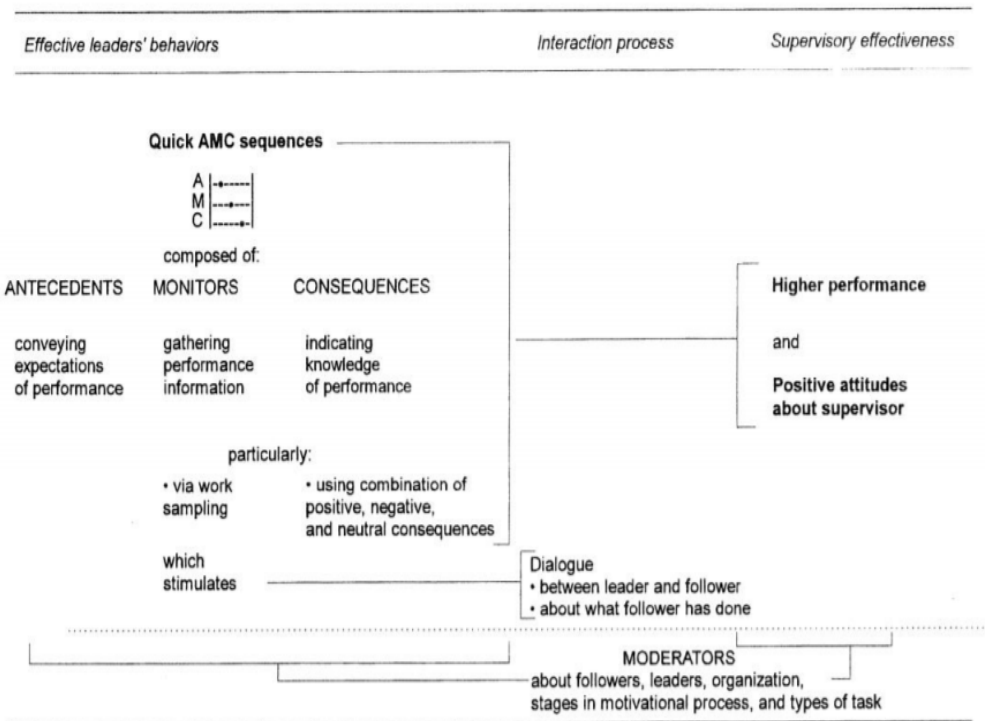


Figure 1. Components of the operant model of effective supervision. Reprinted from *Leadership from an Operant Perspective* (p. 26), by J. L. Komaki, 1998, London: Routledge. Copyright 1998 by Judith L. Komaki. Reprinted with permission.