THE EFFECT OF PERFORMANCE VARIATION ON RATER ATTRIBUTIONS AND RATINGS OF JOB PERFORMANCE

A Dissertation

by

SAURABH SHYAMSUNDER DESHPANDE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Stephanie C. Payne
Committee Members,	Winfred Arthur, Jr.
	Charles Samuelson
	Victor Willson
Head of Department,	Douglas Woods

December 2015

Major Subject: Psychology

ABSTRACT

In most organizations, employee performance is evaluated annually by their supervisors and these evaluations lead to important individual and organizational outcomes. Research has shown that properties of the performance distribution referred to as Gestalt characteristics (e.g., mean, variability) have a significant effect on performance ratings, as well as the attributions that raters make about ratee ability and motivation. This study extends previous research demonstrating the influence of Gestalt characteristics on performance ratings by examining the effect of two operationalizations of variability on performance ratings: tremors (short-term changes) and swells (longer-term changes). One hundred forty-eight participants participated in a 3 (mean: below average, above average, average) × 3 (swells: positive, negative, and none) × 2 (tremors: low, high) × 2 (rater locus of control: internal, external) mixed factorial experiment. Participants evaluated 18 hypothetical salespersons' performance distributions and made attributions about the salesperson's locus of causality, ability, and effort. Findings indicated that both tremors and swells had a significant effect on performance ratings, such that performance profiles with a high level of tremors were rated more favorably than profiles with a low level of tremors, and profiles with a positive/negative swell were rated significantly higher/lower than profiles without a swell. As predicted, tremors had a significant effect on rater attributions of effort such that raters attributed higher amounts of effort to performance profiles with a high level of tremors compared to profiles with a low level of tremors. Swells had a significant effect on rater attributions of ability as well as attributions of effort such that raters attributed positive swells to higher levels of ability and effort and negative swells to lower levels of ability and effort. Contrary to expectation, rater locus of control did not moderate the variability-performance rating relationship. However, exploratory analyses revealed that rater locus of control moderated the relationship between swells and attributions of locus of causality, such that raters with an internal locus of control tended to attribute swells internally rather than externally. Implications of these findings for performance management are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

INTRODUCTION

There is ample evidence that individual work performance varies from time to time and is dynamic in nature (Deadrick, Benett, & Russell, 1997; Deadrick & Madigan, 1990; Ghiselli & Haire, 1960; Hoffman, Jacobs, & Gerras, 1992; Hulin, Henry, & Noon, 1990; Thoresen, Bradley, Bliese, & Thoresen, 2004). Variability in performance occurs because people are not likely to perform “at their best” at all the times for various reasons (Beal, Weiss, Barros, & MacDermid, 2005). Correspondingly, performance is likely to vary within a workday as well as across work days. In fact, research has shown variability *within* person (intraindividual variability) can be significantly larger than variability *between* persons (Fisher & Noble, 2004; Stewart & Nandkeolyar, 2006). In a review of 36 independent samples, Dalal, Bhave, and Fiset (2014) reported that 62% of the variability in task performance could be attributed to within person sources.

Although the influence of intraindividual variability on performance ratings has been empirically examined in multiple studies, research examining variability has primarily focused on point-to-point fluctuations (tremors) or longer-term trends. As a result, researchers have called for an examination of other types of variation, particularly short-term fluctuations, on performance ratings (Reb & Greguras, 2008; Stewart & Nandkeolyar, 2006).

In an effort to explain why raters evaluate variation the way they do, the attributions that raters make about ratees will also be examined. Although attribution theory is used extensively in social psychology research, it has not played a prominent role in the industrial/organizational (I/O) psychology (Martinko, 1995). Martinko, Douglas, and Harvey (2006) noted that the role of attributions is significant for behaviors relevant to central topics in the I/O psychology including performance appraisal. Harvey, Madison, Martinko, Russell, Crook, and Crook (2014) concluded that attributions are an important and integral part of cognitive processes and are associated with critical organizational outcomes frequently based on performance ratings. In this study, attribution theory helps to explain rater cognitions when evaluating performance.

For some occupations, intraindividual variability in job performance can have a significant influence on employee compensation. For example, basketball shooting variability is significantly related to salary for professional basketball players, with a high level of variability being associated with in lower salaries (Barnes & Morgeson, 2007). Similarly, salespersons earning only commission on their sales and not salary may experience significant lapses in compensation when they experience a "dry spell" in which they do not sell anything. Correspondingly, it is important to understand how intraindividual variability influences the evaluation of employees for the following reasons. First, given the prevalence of performance variability, supervisors should be informed of the extent to which performance variability impacts ratings in the first place. Suppose there are two employees with the same level of mean performance. Is the employee with higher variability rated higher or lower than the employee with lower variability? Second, given multiple types of performance variability, it would be informative to ratees to know how raters attribute performance variability, and that they should inform raters when there are extenuating circumstances that contributed to their performance. Supervisors may have difficulty distinguishing between momentary changes and short-term changes. Supervisors may overreact if they interpret momentary changes to be short-term changes. On the other hand, supervisors may overlook an opportunity for positive or negative feedback if they interpret short-term changes as momentary fluctuations.

Given the evidence for intraindividual variability in job performance and its influence on supervisor ratings of performance which are used to make important personnel decisions, it is important to understand how intraindividual variability influences supervisor ratings and the extent to which other variables influence how intraindividual variability is interpreted and evaluated. For example, are there properties of the performance distribution (e.g., mean, swells) that augment or diminish the effect of intraindividual variability on supervisor ratings? Are raters more likely to attribute intraindividual variability to the ratee than external sources? Does rater locus of control influence these attributions? This study begins to answer these questions. Before doing so, various forms of intraindividual variability will be defined.

Intraindividual Variability in Performance

Intraindividual performance variability or within person performance variability is defined as fluctuations in the performance level of an employee over a given period of time (Dalal et al., 2014). Traditional views of performance appraisal completely ignore within person variability either implicitly or explicitly by interpreting it as random error, sampling error, or systematic error (Beal et al., 2005; Dalal et al., 2014; Deming, 1986). Dalal et al. (2014) concluded that researchers have long have neglected intraindividual performance variability or have simply ignored it as measurement error. Although between-person differences have advanced theories of performance and made genuine contributions, within person variability remains understudied and needs due attention (Dalal et al. 2014).

Over the years, there have been attempts to measure and account for intraindividual variability. For instance, Kane (1984, 1986) introduced the performance distribution assessment (PDA) method, which was designed to capture variability in performance. Kane found that intraindividual performance variability can vary quite extensively across individuals. The PDA method requires raters to categorize ratee's performance into pre-determined performance bands (determined by the organization) and assign a frequency count to each band. Using a set of complex calculations, the PDA method provides a final performance rating. However, the PDA method forces the rater to assign the employee to a fixed band and therefore does not result in a fine-grained assessment of performance which is necessary to identify subtle differences between employees. Secondly, the PDA method does not graphically depict performance over time. This temporal mapping can be useful in visualizing and identifying trends, cycles, and patterns of behavior.

Researchers have described and categorized intraindividual performance variability in a number of different ways. Ultimately, intraindividual performance variability can be described with regard to four primary dimensions: (1) linearity: "linear vs. non-linear trends or growth curves"; (2) duration: the amount of time they represent; (3) permanence: permanent changes vs. temporary changes that are reversible; and (4) cycle: "approximate sinusoidal functions of recurring peak and troughs" vs.

discontinuous event-driven sudden changes which form peaks and troughs that are nonrecurring (Dalal et al., 2014, p.1400).

Lumsden (1977) classified intraindividual variability into three types based on duration and permanence: (1) *trends*, (2) *swells*, and (3) *tremors*. Trends are long-lasting permanent changes, swells are short-term fluctuations, and tremors are point-to-point variations or momentary fluctuations (i.e., outliers). Short-term is a relative concept based on the duration of the study. Both swells and tremors are temporary and reversible changes. Nesselroade (2001) identified two types of intraindividual variability based on duration and permanence: (1) *intraindividual change*, representing a long and permanent change and (2) *intraindividual variability* representing short and reversible change. Reb and Cropanzano (2007) similarly differentiated changes in performance profiles based on duration and permanence as: (1) long term changes, referred to as *performance trends,* that reflect a permanent increase or decrease in the mean level of performance over an extended period of time, and (2) short term changes, referred to as *performance variations* around the mean that are temporary and reversible in nature.

Although there are multiple classifications of intraindividual variability, the various types of variability described are overlapping and all seem to take into consideration duration and permanence. In the current study, Lumsden's (1977) terms are used to describe the focal types of variability. In this study, performance is depicted weekly over 26 week period. Thus, *tremors* are week-to-week changes in performance. *Swells* are longer periods of variation operationalized as a deviation from the mean for three weeks. Conceptually, swells have properties similar to outliers in that they represent a brief deviation from the mean, but they also represent a form of short-term variability. Although variability in performance is found in various occupations, it is particularly prevalent in the sales industry and therefore serves as a relevant context for this vignette-based study.

Intraindividual Variability in Sales

Approximately, 4.2 million people are employed in retail sales and about 1.86 million people are employed in technical sales (US Department of Labor, Bureau of Labor Statistics, 2014). Recent studies have shown that variability within salespersons is greater than variability across salespersons (Stewart & Nandkeolyar, 2006). Stewart and Nandkeolyar collected data for 167 sales representatives across 26 weeks and found that only 27% of the variance could be attributed to between salesperson differences and 73% of the variability could be attributed to within salesperson differences. These data suggest that there is a high level of intraindividual variability in the sales profession.

In the sales industry, negative swells occur when salespersons who typically perform well experience a short term decline in or lack of sales and are colloquially referred to as “dry spells” or “slumps.” Writing in the popular press suggests that slumps are prevalent and attempts to rectify them by proposing various strategies on how to get out of a slump. For example, a simple search on the Internet yields a myriad of results for articles on how to manage a dry spell or slump. A search of a popular bookstore website (Barnes & Noble) for “sales slump” yielded four different books exclusively on how to deal with a slump in sales (*There's a Fine Line between a Groove and a Rut: How to Avoid a Sales Slump and Re-energize Your Marketing Team, August 2005,* by G.

D. Kittredge III; *Sales Slumps: Pinpoint Sales Management Skill Development Training Series, April 2011,* by Timothy F. Bednarz, Majorium Business Press; *The Sales Slump Doctor Is In!, February 2001,* by Mickey M. Greenfield; and, *Bust Your Slump: A Dozen Strategies to Fill Your Pipeline in 30 Days, August 2010,* by Paul M. McCord).

In contrast to the popular press, there is a relatively limited amount of peer-reviewed research on sales slumps. A search for "sales slumps" in PsycINFO for the years of 1970 to 2014 yielded a single article by Connell titled "What to do when selling slump hits" published in 1977 in the *Training and Development Journal* which describes the use of transactional analysis for the training of salespersons when they hit a slump. A similar search in the *Journal of Personal Selling & Sales Management* yielded a single article by Wortuba and Schoel titled "Evaluation of Salesforce Contest Performance" published in 1983 that focused on various aspects of a sales contest. Furthermore, a search through the journal of *Sales and Marketing Management* yielded 67 articles including topics on sales slumps, profit slumps, economy slumps, and suggestions on what to do in a sales slump. However, no studies were found examining the impact of slumps on the performance ratings of salespersons.

In the following sections, Gestalt characteristics are introduced, followed by a brief review of the cognitive process of rating, attribution theories, and finally the effects of Gestalt characteristics on rater attribution are put forth.

Gestalt Characteristics

Supervisors face a challenging cognitive task when they are asked to provide summary ratings of variable job performance. They have to process a relatively large amount of information, often over an extended period of time, and then frequently aggregate the ratings to arrive at a single score (Reb & Cropanzano, 2007). For many supervisors, they must do this for multiple direct reports. Furthermore, due to cognitive resource limitations and the absence of aggregation rules, supervisors may lack the motivation or capability to process the large amount of data optimally (Reb & Cropanzano, 2007). The more information they have to consider, the more complex the rating process is for them. This process is further complicated when supervisors are asked to make between-employee comparisons, which is one of the primary reasons for evaluating performance (e.g., Cleveland, Murphy, & Williams, 1989). Not only does intraindividual variability make rating a challenging task, but it also introduces more opportunities for raters to diverge in their ratings, a well-established phenomenon in the 360-degree performance evaluation research literature (Harris & Schaubroeck, 1988; Viswesvaran, Schmidt, & Ones, 2002) as different raters may weigh the variability in the performance profiles differently.

One way to address the cognitive limitations of raters is to gather and depict performance data on a graph. This technique is relatively common when there are objective data gathered regularly such as sales as they provide a lot of information to the rater in a single glance (Williams, 2014). Recent studies on intraindividual variability (Lee & Dalal, 2011; Reb & Cropanzano, 2007; Reb & Greguras, 2010) plot performance data (e.g., sales dollars) on a line graph relative to time. These graphs are sometimes referred to as temporal performance profiles as they depict multiple data points representing quantity or quality of work-related behavior over a specified time period.

Often, they also provide normative information such as company averages or standards for employees doing the same work.

Research has demonstrated that people use salient structures of observed data depicted in a graph to generate a performance rating. These salient, attention-getting features of the performance profile have been referred to as “Gestalt characteristics” that help people to organize the data points to arrive at a holistic perception (Koffka, 1935) and are believed to drive holistic evaluations (Ariely & Carmon, 2000, 2003). These characteristics provide an efficient representation of the data leading to increased memory efficiency helping to overcome human cognitive capacity and motivation limitations (Ariely & Carmon, 2003; Gilovich, Griffin, & Kahneman, 2002). Correspondingly, supervisors are likely to focus on Gestalt characteristics of the performance distribution as this requires less effort than thinking about each data point separately (Gilovich et al., 2002; Kahneman, Slovic, & Tversky, 1982).

A profile of data points has three main characteristics that can be inferred from the pictorial representation of data: (a) mean level of performance or the mathematical average of the data depicted on the graph, (b) trends, long lasting upward or downward changes, and (c) variation or tremors, point-to-point fluctuations (Reb & Cropanzano, 2007; Reb & Greguras, 2010). Other characteristics of the temporal profile that are likely to be considered because of their salience include peaks and troughs/valleys (high and low extremities) and the endpoints (Ariely & Carmon, 2003; Kahneman, 2000). Extremities, or extreme scores, in a profile stand out because of the magnitude of their departure from the mean thereby making them salient points as demonstrated by Lee and Dalal (2011).

Similarly increasing or decreasing trends are also likely to stand out because they demonstrate an identifiable pattern within the data (Lee & Dalal, 2011; Reb & Cropanzano, 2007; Reb & Greguras, 2010).

A number of studies have examined the influence of Gestalt characteristics of job performance profiles on performance ratings (DeNisi & Stevens, 1981; Lee & Dalal, 2011, Reb & Cropanzano, 2007; Reb & Greguras, 2010). In these studies, researchers manipulate Gestalt characteristics of hypothetical performance distributions displayed in graphical form. Typically, a group of undergraduate students are presented with a set of graphs and are tasked with rating performance and sometimes making judgments about the ratee's ability and motivation. Earlier studies by Scott and Hamner (1975) and DeNisi and Stevens (1981) and relatively more recent studies by Reb and Cropanzano (2007), Reb and Greguras (2010), and Lee and Dalal (2011) have provided some insights regarding how Gestalt performance characteristics influence performance ratings.

In the next section, a chronological review of the studies examining Gestalt characteristics in performance profiles is provided. Additionally, a summary of the findings with regard to the three most studied characteristics (mean, variation, and trend) is provided in Table 1. Various attribution theories are used to explain why Gestalt characteristics influence ratings the way they do.

Summary of Gestalt Characteristic Studies

Scott and Hamner (1975) investigated the influence of variations and trends in performance profiles on supervisory ratings of workers' ability, motivation, and overall worker performance using sixty male business students in a laboratory experiment. They conducted a 2 (variation: high, low) × 3 (trend: ascending, random, descending) factorial design. Contrary to expectation, the results of their study indicated that neither variation nor trend affected overall performance ratings in any way. Scott and Hamner (1975) speculated that the raters may have thought the variation and trend were random error. The raters may not have even detected the trends as they evaluated the workers while they were performing, the raters were not provided with graphical profiles of performance for easy trend identification, and a manipulation check was not conducted. It cannot be determined if the lack of significant results for variability and trend were due to the absence of a graphical depiction of the data or a weak manipulation.

DeNisi and Stevens (1981) manipulated the average level of performance, variation, and trend. They conducted a 2 (variation: stable, variable) × 3 (mean: high, average, low) × 2 (trend: ascending, descending) experiment with 147 undergraduate student raters. Their results indicated that the average level of performance, or mean, was found to be the most important determinant of performance ratings. Similar to Scott and Hamner's (1975) results, variation did not have a significant main effect on ratings. However, mean and variation interacted with one another to predict performance ratings. For low and average levels of performance, stable performance received significantly higher ratings than variable performance. In contrast, variation did not have a significant effect when performance was high.

Reb and Cropanzano (2007) presented 35 subordinate profiles (five of which were repeated for test-retest purposes) to a group of 64 undergraduate student raters. They conducted a more complex experiment with a 5 (trend: flat, ascending, descending, U-shaped, ∩-shaped) × 3 (mean relative to a norm: negative, zero, positive) × 2 (variation: small, large) factorial design. These within-subject manipulations yielded 30 unique profiles. They also manipulated display factor (graphical vs. tabular) as a between-person variable. The results of their study revealed that compared to a flat trend, an ascending trend led to more positive ratings, whereas a descending trend led to lower ratings. Higher means led to substantially higher ratings of performance, whereas, lower means led to lower ratings in both graphical as well as tabular displays. Performance variation did not, however, have a significant effect on ratings. The authors speculated that the manipulation of variation may not have been strong enough.

Reb and Greguras (2010) manipulated characteristics of performance distributions in a 3 (mean: below average, average, above average) × 3 (trend: ascending, flat, deteriorating), × 2 (variation: small, large) factorial design. They investigated the interaction of the purpose of the ratings (administrative vs. developmental) with mean, trend, and variation. Their results for trend and mean were similar to Reb and Cropanzano's (2007) findings. However, variation played a significant role in performance ratings such that more variation led to less favorable ratings, which they attributed to a stronger manipulation of variation. (They doubled the value of the standard deviation used in the Reb and Cropanzano's study.)

A recent study by Lee and Dalal (2011) focused specifically on performance extremities which are frequently called outliers in statistics. They examined the influence of three levels of extremities: none, one, or two. It is important to note that in

Table 1

Summary of the Gestalt Characteristics Studies

Study	Mean	Variation	Trend[a]	Interaction between Mean and Variation
Scott & Hamner (1975)	Not manipulated	Not significant	Not significant	
DeNisi & Stevens (1981)	Positive main effect	Not significant	Main effect	For below average and average levels of performance, stable performance was rated higher than variable performance
Reb & Cropanzano (2007)	Positive main effect	Not significant	Main effect	
Reb & Greguras (2010)	Positive main effect	Negative main effect	Main effect	For below average performance, stable performance was rated higher than variable performance
Lee & Dalal (2011)	Positive main effect	Not manipulated	Main effect	

Note. [a]In all studies examining trend, a descending trend had a negative effect and an ascending trend had a positive effect.

the two extremity conditions, the two extreme scores were not depicted consecutively. They replicated previous effects of mean and trend on performance ratings (DeNisi & Stevens, 1981; Reb & Cropanzano, 2007; Reb & Greguras, 2010). Negative extreme scores had a significant negative effect on performance ratings but the same effect was not observed for positive extremities. Thus, negative extremities had a stronger effect on performance ratings than positive extremities. It appears that raters penalize ratees when negative extreme scores are present.

This current study seeks to replicate Reb and Greguras's (2010) finding that variation is negatively related to performance ratings. The current study also extends Lee and Dalal's (2011) study on outliers, by examining the influence of a swell, a cluster of consecutive outliers, on ratings of performance. In summary, studies of Gestalt characteristics have shown that the dynamic characteristics of performance profiles such as the mean, trend, variation, and outliers have a significant effect on performance ratings. Results are fairly consistent for mean and trend such that higher mean levels and positive trends result in higher ratings of performance. Results for variation tend to be nonsignificant but arguably inconclusive.

The present study examines two operationalizations of variability in a performance distribution: tremors and swells. Tremors are operationalized as the standard deviation of the performance data. With regard to swells, there does not appear to be a universally agreed upon way of operationalizing a short-term fluctuation in a performance profile. In contrast, when the cluster appears below the mean, it is referred to as a negative swell. In order to provide a holistic picture of the rating process, a review of the cognitive processes involved in evaluation is described next.

Cognitive Process of Rating

A number of performance appraisal researchers have proposed models that depict the cognitive processes involved in evaluating others based on the social information processing theory (DeNisi, Cafferty, & Meligno, 1984; Feldman, 1981; Murphy & Cleveland, 1995; Salancik & Pfeffer, 1978). All of these models illustrate multiple stages or steps that a rater experiences. First, raters observe an ideally representative sample of behavior (Green & Mitchell, 1979) and sometimes organizational records of performance. Some of this information is encoded and stored into long-term memory (DeNisi et al., 1984; Feldman, 1981; Ilgen & Feldman, 1983). When it is time to complete an evaluation, the performance-related information from various sources needs to be retrieved, integrated, and evaluated in order to generate a rating.

Feldman (1981) proposed that when evaluating others, people are inclined to engage in an automatic categorization of information. These categories may be based on sex, age, race, and appearance of the ratee or they can be based on traits (introvert, extrovert, neurotic, friendly) that the rater may use regularly to categorize people. However, a controlled, and therefore less biased, categorization process is more likely to happen when substantial information about the ratee is not available for the rater to perform automatic categorization (Feldman, 1981). The first stage in this controlled process is the causal attribution. This is when the rater determines whether the ratee's behavior is caused by the ratee or by situational factors external to the ratee. Assumptions about causes of behavior are called attributions. Attributions are important to the assessment of performance as they determine how much the rater holds the ratee accountable for his/her behavior and directly impact the rater's evaluation. In the next section, attribution theories relevant to the study are reviewed, and empirical research examining supervisor attributions of subordinate performance is summarized.

ATTRIBUTION THEORY

Heider (1958) proposed that people are "naïve" psychologists trying to identify cause and effect relationships in order to make sense of the social world. When they identify potential causes of behavior, they attribute the behavior to, or regard it as resulting from, one or more causes. Attribution theories explain how people make causal inferences, what information they use, and how they arrive at these inferences on the basis of this information (Kelley, 1972). Among the numerous attribution theories, the ones most relevant to supervisor attributions of ratee performance are (1) Kelley's (1972) covariation model which provides predictions about how raters will attribute performance, (2) Weiner's (1971, 1974, 1985) model of achievement attributions which identifies the four determinants of achievement based on stability and locus of causality, and (3) Ross's (1977) fundamental attribution error which proposes that raters downplay or ignore situational factors, and (4) Gilbert and Malone's (1995) correspondence bias which predicts that raters tend to attribute behavior of others to the ratee's disposition expanding on the fundamental attribution error.

Kelley's (1972) Covariation Model: Dispositional vs. Situational Causes for Behavior

One of the most basic distinctions that people can make when observing someone else's behavior is to attribute behavior to internal or external causes (Heider, 1958; Weiner, Frieze, Kukla, Reed, Rest & Rosenbaum, 1971). Internal causes are traditionally dispositional characteristics (personality, character, ability, and other stable traits) and external causes are situational factors (the environment or situation such as task difficulty, market conditions, and luck; Heider, 1958). Correspondingly, a fundamental dimension on which to evaluate attributions is the *locus of causality*: internal vs. external.

According to Kelley's (1972, 1973) covariation model, a rater considers three pieces of information when making an internal vs. an external attribution. It is assumed that the rater has performance information from multiple observations at various times and situations such that the rater can perceive covariation in performance and the causes for the variance. The three pieces of information are: (a) *consensus*: how other people behave in a similar situation (between-person variability); (b) *distinctiveness*: if the ratee behaves the same way in different situations (tasks; within-person variability across tasks); and, (c) *consistency*: if the ratee repeats the behavior every time the situation occurs (task is performed; within-person variability on the same task). According to Kelley, raters make internal attributions when (1) consensus is low (other people do not behave the same way), (2) distinctiveness is low (the ratee's behavior is not unique to the situation/task at hand) and (3) consistency is high (the ratee behaves the same way most of the time). On the other hand, raters make external attributions when consensus is high, distinctiveness is high, and consistency is low.

In the present study, raters are presented with 18 different performance profiles representing 18 hypothetical salespersons. With regard to Kelley's (1972) key pieces of information, consensus in this study is conveyed by the similarity of performance profiles across salespersons. This is intentionally manipulated such that no two profiles are exactly the same, making other people's behavior variable and consensus conceivably low. Hypothetical ratees are performing the same task every week; thus there is no opportunity for raters to evaluate ratees in different situations, making behavior on the task indistinct. Thus, distinctiveness is constant and unlikely to contribute to rater attributions. With regard to the final piece of information in Kelley's model, the consistency of the ratees' performance is also manipulated by altering within person variation over the course of 26 weeks, so consistency should be perceived as low. In summary, raters are asked to make attributions about performance when consensus is low, there is no information about distinctiveness, and consistency is low. Again, low consensus is associated with internal attributions but low consistency is associated with external attributions, so there are conflicting pieces of information.

When competing causes are present, Kelley's (1972) discounting principle states the rater will attribute the behavior to one of the causes to a lesser magnitude (discount that cause) than if the rater was aware of a single cause. Kelley further postulated that when sufficient information is not available, the rater relies on his/her personal experiences to arrive at a causal inference.

Attribution Errors and Biases

Kelley (1972, 1973) noted that although attribution principles are very sensible and rational ways to draw inferences, people engage in less rational attributional tendencies which are called biases or errors. According to Kelley (1972), "Too little account is taken of external causes (contextual factors) in judgments of other person's behavior" (p. 18). This tendency was later referred to as the *fundamental attribution error*, by Lee Ross (1977), which occurs when a rater overemphasizes the dispositional characteristics of the ratee and downplays or ignores situational characteristics. Based on the fundamental attribution error, raters are more likely to attribute other people's performance internally rather than externally.

Extending the work on attribution error, Gilbert and Malone (1995) introduced the concept of correspondence bias. Correspondence bias occurs when a rater observing someone else's behavior attributes the behavior to the ratee's unique dispositions and concludes that the ratee has a predisposition for that behavior, especially when those behaviors cannot be explained by situational factors. According to Gilbert and Malone (1995), correspondence bias occurs for four main reasons: "(a) raters lack awareness of the ratee's situation as it is objectively constituted or subjectively construed, (b) the rater has inappropriate expectations for how a person will behave in such a situation, (c) the rater's awareness of the ratee's situation has led to an inaccurate perception of the ratee's behavior, or (d) the rater lacks either the motivation or the capacity to correct the trait inferences he or she may have spontaneously and effortlessly made" (p. 30). In summary, Gilbert and Malone extended the impact of the fundamental attribution error to attributions about others' dispositions and identified four mechanisms that facilitate the error and corresponding bias.

Weiner's (1971, 1974, 1985) Achievement Attributions

Attribution theorists have gone well beyond the internal vs. external dichotomy (locus of causality dimension) when classifying attributions. Another core dimension on which attributions vary is the stability of the cause: stable or unstable. When locus of causality and stability are crossed, performance can be attributed to one of four sources:

ability, effort, luck, and task difficulty (Weiner et al., 1971). As depicted in Table 2, ability and effort are internal, whereas luck and task difficulty are external. Further, ability and task difficulty are stable, whereas effort and luck are unstable. Stable performance is likely to be attributed to stable determinants (ability or task difficulty). In contrast, unstable (variable) performance is likely to be attributed to unstable determinants (effort or luck) (Weiner et al., 1971).

Table 2

Classification Scheme for the Perceived Determinants of Achievement Behavior

Stability	Locus of Causality	
	Internal	External
Stable	Ability	Task Difficulty
Unstable	Effort	Luck

Note. Adapted from Weiner et al. (1971, p. 96) and Weiner (1985).

In the current study, the task performed by the hypothetical employee does not change; thus, task difficulty remains the same. Therefore, it is unlikely that a rater would attribute performance to task difficulty. This leaves three possible causes to which raters can attribute performance (ability, effort, and luck).

Later, Weiner (1985) acknowledged that the locus of control dimension was frequently confused with Rotter's (1966) individual difference variable called "locus of control." Weiner proposed changing the label of his dimension to *locus of causality* to refer to the extent to which performance is attributed to internal vs. external causes. Weiner described performance as a function of ability, effort, task difficulty, and luck but did not propose a specific mathematical formula for how to combine these variables to predict attributions.

Weiner's (1971, 1974, 1985) model has been tested and refined numerous times. In most attribution studies, researchers examined participants' attributions for their own success or failure (Bar-Tal, Goldberg & Knaani, 1984; Elig &Frieze, 1979; Wilson & Palmer, 1983), but they sometimes examined attributions made about others' success and failure as well (Anderson, 1983; Burger, Cooper & Good, 1982; Cooper & Burger, 1980; Frieze & Weiner, 1971). All of these studies reveal that "success is ascribed to high ability and hard work, and failure is attributed to low ability and the absence of trying" (Weiner, 1985, p. 549). Frieze and Weiner (1971) manipulated Kelley's attributional criteria to investigate causal judgments by raters using all of the four causal factors in Weiner's model. They manipulated consistency (operationalized in this study as percentage of success for a hypothetical person at a particular task), distinctiveness (percentage of success of that same person on other tasks), and consensus (percentage of other individuals' success at the same task). The participants were then asked to judge the two outcomes (success and failure) to be attributed to ability, effort, task difficulty, and luck. They found that the raters attributed inconsistent performance (ratee with a history of success has a failure or vice versa) to unstable causes, luck, and effort. When the performance was consistent (success followed by success and failure followed by failure), the raters attributed it to stable causes, task difficulty, and ability. These results suggest that raters may be more likely to attribute variability in performance to unstable causes like luck and effort than stable causes, but it does not reveal if there is a tendency to attribute variability more so to internal or external causes.

CURRENT STUDY

In contrast to the traditional attribution studies where participants are asked to make judgments or ratings about a single event or outcome of behavior (often construed as success or failure); in the current study, raters are asked to evaluate multiple data points in a single profile representing performance of a hypothetical employee over time. Furthermore, the performance outcomes in the traditional attribution studies were typically dichotomous (success or failure), whereas in the current study, raters are provided performance outcome data on a continuous scale. Instead of observing behavior, raters examine a distribution of data points on a line graph. Raters are asked to judge the entire performance profile consisting of 26 data points plotted on a graph and make attributions about ability, effort, and locus of causality.

In the current study, information about consistency and consensus are indirectly altered by manipulating the Gestalt characteristics (variation and swells) of multiple performance profiles. How a rater evaluates a given temporal profile is likely to be a function of the attributions he/she makes about what causes or contributes to the Gestalt characteristics of the performance profile, as well as his or her beliefs, tendencies, and biases.

THE INFLUENCE OF GESTALT CHARACTERISTICS ON RATER ATTRIBUTIONS

Performance can be attributed to ability, effort, task difficulty, and luck/situational factors (e.g., market conditions; Weiner et al., 1971). In an effort to understand the previously discussed effects of Gestalt characteristics on performance ratings and rater attributions, theory and research are used to develop the hypotheses presented below.

Performance Mean

As previously noted, the mean of a performance distribution is simply the arithmetic average of the data points depicted for a given period of time. Organizations sometimes use set norms or long time averages (across multiple salespersons and multiple years) as relative information for the rater to consider when evaluating the performance of an individual salesperson. This standard is often depicted on a temporal graph of performance, making it easy for the rater to see if the employee's performance is consistently above, at, or below the organization's mean.

Deming (1986) speculated that supervisors fail to discriminate between person and system variables that lead to poor employee performance, and they tend to believe that poor performance is caused exclusively by variables within the employee, ignoring system factors. Research has shown that raters are more likely to attribute performance to the ratee especially when performance is poor (Carson, Cardy & Dobbins, 1991; Fedor & Rowland, 1989; Jones & Nisbett, 1972; Knowlton & Mitchell, 1980; Mitchell & Wood, 1980; Ross, 1977). For example, Carson et al. (1991) found that raters ignored system factors and blamed ratees for poor performance in spite of knowing that system factors were the cause of poor performance. Fedor and Rowland (1989) also posited that raters hold subordinates responsible for poor performance levels which may be due to factors beyond their control. They explained that such bias in judgments happens as raters are insensitive to external constraints and difficulties (Jones & Nisbett, 1972). When no information is provided about the situation, it is even more likely that raters will attribute low performance to the ratee. Thus, based on the fundamental attribution error and correspondence bias and previous empirical research, below average performance is expected to be attributed internally or to being within the ratee's control.

H1: There will be a main effect for mean level of performance on attributions of locus of causality such that below average levels of performance are more likely to be attributed internally than externally.

Numerous studies have demonstrated that cognitive ability is a strong predictor of performance across jobs; thus it explains mean differences in job performance (e.g., Hunter & Hunter, 1984). For a given task, perceived ability is inferred based on past success on similar tasks (Weiner et al., 1972). In a performance profile, unless a performance standard is explicitly provided by management, success may be represented by above average performance and failure may be represented by below average performance. Raters are likely to attribute above average performance to a high level of ability and empirical studies of performance profiles have demonstrated this effect (e.g., Reb & Greguras, 2010).

Ability alone, however, is not the only variable that influences performance.

Recognizing the influence of motivation on performance, raters are also likely to attribute performance to effort expended. Higher effort levels will lead to higher levels of performance; thus raters are also likely to attribute above average performance to higher levels of effort (Reb & Greguras, 2010; Russell, McAuley, & Tarico, 1987).

H2: There will be a positive main effect for mean level of performance on judged ability.

H3: There will be a positive main effect for mean level of performance on judged effort.

As previously noted, empirical research has consistently demonstrated that the mean of a performance profile has a significant effect on performance ratings, with higher levels resulting in higher ratings. In fact, the majority of the variance in performance ratings can be explained by the mean (DeNisi & Stevens, 1981; Lee & Dalal, 2011; Reb & Cropanazano, 2007; Reb & Greguras, 2010, Scott & Hamner, 1975). According to DeNisi and Stevens (1981), "the level of performance should be a major determinant of the overall ratings" (p. 594). Correspondingly, a main effect for mean is expected on performance.

H4: There will be a positive main effect for the mean level of performance on performance ratings.

Performance Variation

Another important variable to consider when evaluating performance over time is variation. Variability occurs both within person and between persons. This study focuses on within-person variability. Thus, variation represents how much an individual's performance deviates from his/her average level of performance within the time span of interest. Both short-term (tremors) and longer-term (swells) variation are examined in this study.

Tremors

Short-term, point-to-point fluctuations are referred to as tremors (Lumsden, 1977). This type of variation is operationalized statistically as the variance or the standard deviation of the performance distribution. A performance profile with a low level of tremors will have most data points appearing relatively close to the average level of performance; whereas a performance profile with a high level of tremors will have multiple data points appearing further away from the average.

As noted earlier, one of the key variables that determines rater attributions is consistency in behavior or performance (Kelley, 1972, 1973). Consistency is the opposite of variability. According to Kelley's model, low consistency (within-person variability) is expected to be attributed externally. However, Kelley's model also identifies between-person variability (consensus) as an important variable to consider when making attributions. When there is a high amount of between-person variability in performance (low consensus), the rater is likely to attribute the focal employee's performance internally. Thus, Kelley's model does not provide a clear prediction on how a rater will attribute within-person variability in a context of between-person variability (low consistency and low consensus).

Unfortunately, Weiner's (1972) model does not provide a clear prediction for the influence of variation on rater attributions either. According to Weiner's model, variable performance is likely to be attributed to unstable determinants. These could be effort which is perceived as internal or luck which is external. Thus, raters could attribute within-person variability internally or externally.

As noted earlier, extensive research documenting the fundamental attribution error indicates that raters are likely to attribute others' performance internally, regardless of whether it is consistent or not. This would suggest that raters would attribute others' within-person variability internally. On the other hand, Reb and Greguras (2010) examined rater attributions for within-person variability in the context of between-person variability and found that raters tended to attribute variable performance externally. Due to the competing predictions offered by theory and limited empirical research on the influence of variability on locus of causality, the influence of performance variation on locus of causality is therefore framed as a research question.

RQ 1: What is the effect of tremors on attributions of locus of causality?

Ashkanasay (1989) asked subjects to read descriptions of performance which varied in consistency, consensus, and distinctiveness and asked them to attribute the performance outcome to ability, task difficulty, luck, and effort. Results revealed that consistent outcomes (success followed by success, or failure followed by failure) were more likely to be attributed to stable causes such as ability and task difficulty than inconsistent outcomes, and inconsistent outcomes (success followed by failure or vice versa) were more likely to be attributed to unstable causes such as luck and effort than consistent outcomes. Correspondingly, variable performance is more likely to be attributed to unstable causes (effort and luck).

The fundamental attribution error states that raters are likely to ignore or downplay situational factors (e.g., luck, market conditions) and therefore tremors are more likely to be attributed internally to effort rather than externally to luck/situational factors (e.g., market conditions).

H5: There will be a positive main effect for tremors on judged effort.

In contrast to the consistent results revealed for the effect of the mean on ratings of performance, effects for variation have not been consistent. For instance, some studies have reported no main effect for variation (Reb & Cropanzano, 2007; Scott & Hammer, 1975). On the other hand, Reb and Greguras (2010) found that variation resulted in higher performance ratings. They attributed this finding to using a stronger manipulation of variation, doubling the standard deviation of variation used in Reb and Cropanzano's (2007) study.

The amount of variation in performance that is tolerated and perceived as acceptable by a supervisor likely depends on industry and organizational norms and the consequences associated with this variability. For example, variation during a coordinated military effort such as a surprise attack on a target may give away the element of surprise and is unlikely to be acceptable or tolerated. Similarly, inconsistency in sentencing for similar offences by a judge or variation in grading for similar answers by instructors is not likely to be perceived as fair and therefore unacceptable. On the other hand, variation in sales is very common (Stewart & Nandkeolyar, 2006), and the consequences of this variability are not likely to be life-threatening to anyone. Therefore, variability in sales performance is likely to be expected and acceptable.

How much variability in sales performance is acceptable? This is difficult to answer. Comparing the results of Reb and Cropanzano's (2007) study to Reb and Gregarus's (2010) study provides some potential insight. A small amount of variability (a low level of tremors) did not have a significant effect on ratings (Reb & Cropanzano, 2007), but a large amount of variability (a high level of tremors) did have a negative impact on ratings (Reb & Gregarus, 2010). Due to the inconsistent results, the effect of variation on performance ratings is posed as a research question.

RQ2: What is the effect of tremors on ratings of performance?

Swells

Another form of variation that may influence a rater's assessment of performance is a swell in the performance distribution or short-term fluctuation. These surges are similar to outliers in that they depict a departure from the mean. When swells appear above the mean, they are labelled positive swells. When they appear below the mean, they are labelled negative swells. At least one study of performance profiles is particularly relevant to predictions about positive and negative swells.

Lee and Dalal (2011) put forth that the departure of a large magnitude from the mean makes an outlier salient and therefore impacts the decision making of the rater in a positive or negative way depending on the direction of the outlier relative to the mean. They examined the influence of one or two positive or negative (but not consecutive) outliers. They did not find a significant increase in performance ratings due to a single positive outlier but they did find a significant decrease in ratings due to a single negative outlier. Lee and Dalal (2011) also found that two negative outliers had an even stronger impact and led to lower performance ratings than one negative point, but they did not find the reverse effect for two positive points. Based on Lee and Dalal's study, it would seem that the presence of negative outliers is associated with lower ratings and this trend appears to be linear such that the more number of negative outliers the lower the rating.

Lee and Dalal (2011) explained their differences in results for positive and negative outliers based on the theory of negativity bias by Rozin and Royman (2001). Rozin and Royman (2001) put forth that people give more weight to negative entities including events, objects, and personal traits than positive ones. Specifically, they noted that "negative events are more salient, potent, dominant in combinations, and generally more efficacious than positive events" (p. 297). Similar conclusions about the stronger effects of negative events are provided by other studies on negative events and negative information (Fiske, 1980; Skowronski & Carlston, 1987, 1989).

As mentioned earlier, raters can easily overlook external constraints and difficulties and may be more likely to attribute performance internally than externally especially when performance is poor (Fedor & Rowland, 1989; Jones & Nesbitt, 1972; Knowlton & Mitchell, 1980). Therefore, a possible explanation for Lee and Dalal's (2011) findings concerning the positive outliers may be that the raters attributed a single positive outlier to external conditions such as luck and something not within the control of the ratee. Similarly, since the two positive outliers used by Lee and Dalal (2011) were not consecutive, they might also be attributed to luck. However, a consecutive cluster of outliers (i.e., a swell) may more likely to be attributed internally to the ratee rather than externally as three or more consecutive points has some predictability (albeit for a shorter period of time). A larger number of points in the cluster could be interpreted as a trend (irreversible, long-term fluctuation) rather than a swell (reversible short-term fluctuation). Based on the fundamental attribution error, Rozin and Royman's (2001) theory of negativity bias, and Lee and Dalal's (2011) findings, raters are likely to hold ratees responsible for poor performance. Thus, negative swells are expected to be attributed internally and to have a stronger impact than positive swells.

H6: There will be a main effect for swells on attributions of locus of causality such that compared to positive swells, negative swells are more likely to be attributed internally.

It is proposed that the placement of consecutive points relative to the mean will have an additive effect such that three (or more) consecutive points above the mean will lead to higher ratings, whereas, three consecutive points below the mean will lead to lower ratings. They also indicate the possibility that the ratee is not performing to his/her full potential.

H7: The presence of a positive swell will lead to higher ratings of (a) performance, (b) judged ability, and (c) judged effort compared to profiles without swells.

H8: The presence of a negative swell will lead to lower ratings of (a) performance, (b) judged ability, and (c) judged effort compared to profiles without swells.

Performance Variation Interactions

The effect of performance variation on ratings may depend on a number of other variables, including other Gestalt characteristics such as the mean and extreme scores. First, the effect of performance variation may depend on the overall level of the performance distribution, such that performance variation depends on whether the overall level of performance is above or below the mean. Previous research has demonstrated an interaction between performance variability and mean level of performance. For example, DeNisi and Stevens (1981) found that for below average and average levels of performance, raters gave higher ratings for stable performance than variable performance. Likewise, Reb and Greguras (2010) found that for below average performance, raters gave higher ratings for consistent performance than inconsistent performance. Thus, it appears that variation is perceived negatively when paired with below average performance. Reb and Greguras (2010) noted that the magnitude of the low data points during periods of variability were lower than any data points in performance distributions with less variability and this likely contributed to the ratings as well.

H9: There will be an interaction between tremors and mean such that a high level of tremors will have a positive effect on ratings when performance is above average, whereas a high level of tremors will have a negative effect on ratings when performance is below average.

As noted earlier, variation can also be operationalized with swells or clusters of outliers. In a review of the literature, no studies examining two operationalizations of variability (swells and tremors) were located. The influence of traditionally operationalized variability on ratings (i.e., tremors) may depend on a second form of variability, swells, such that performance variation paired with positive swells would be rated positively and variation with negative swells would be rated negatively. Consistent with this speculation, Cardy and Dobbins (1994) proposed that raters may infer high potential based on multiple high performance episodes even if they are inconsistent, and assign higher ratings to such distributions of performance. Likewise, raters may focus on the low performance episodes and assign lower ratings as shown in Reb and Greguras's (2010) study.

H10: There will be an interaction between tremors and swells on performance ratings such that a high level of tremors will have a positive effect on ratings when combined with (a) positive swells and a negative effect on ratings when combined with (b) negative swells compared to profiles without swells.

Rater Locus of Control and Causal Attributions

Locus of control is an individual difference variable that is defined as " . . . a generalized expectancy for internal as opposed to external control of reinforcements" (Lefcourt, 1976, p. 27). People with a high internal locus of control are referred to as "internals" whereas people with a high external locus of control are referred to as "externals." Studies have shown that rater locus of control affects the attributions that one makes about his or her own performance (Gilmor & Minton, 1974; Gilmor & Reid, 1979; Lefcourt, Hogg, Struthers, & Holmes, 1975). Performance appraisal researchers have speculated that raters with a high internal locus of control may be more inclined to make internal attributions than raters with an external locus of control (Murphy & Cleveland, 1995). There clearly exists a parallel between the concepts of locus of control and locus of causality such that "an internal locus of control seems to also imply an internal locus of causality" (Martinko & Douglas, 1999; Martinko et al., 2006; p. 150). This suggests that internal raters are more likely to make internal rather than external attributions; however, Martinko (2006) also noted that they were "not aware of any studies that directly investigate the relationship between locus of control and locus of causality" (p. 151). Thus, rater locus of control may also influence the attributions that the rater makes when evaluating others' performance.

Whereas a person's locus of control can be assessed before the event happens, attributions are made after the event has taken place, that is, when the person knows the outcome or effect and is searching for the cause. Reid and Ware (1973) conducted a study to determine if locus of control was transferable to others. They used a videotape of a student discussing his academic failures. Their results showed that subjects who were internals were more likely to blame the student for the failures (internally) than luck or the system (externally). Thus, rater locus of control may affect the raters' attributions regarding locus of causality such that internal raters are more likely attribute performance internally (e.g., to effort) and external raters are more likely to attribute performance externally (e.g., to luck). When a supervisor observes variation in performance, the supervisor may attribute the variation internally to effort or externally to luck or situational factors. Cardy and Dobbins (1994) proposed that raters may infer high potential based on multiple high performance episodes, although inconsistent, and assign higher ratings. It is also possible that the rater may attribute the high performance to "extra effort."

The attribution-related errors/biases are particularly relevant when supervisors make evaluations of performance and assign ratings. When sufficient information is not available, the rater relies on personal information (considering what they personally would have done in that situation or other available historical information; Kelley, 1972). Similarly, when a rater is faced with multiple causes for another person's performance, the discounting principle (Kelley, 1972) is likely to play out. Especially in the situation where a rater can attribute good/bad performance to multiple causes (both internal and external), the rater may downplay the magnitude of the effect of one cause when the rater is aware of other plausible causes. A study by Hansen and Lowe (1976) manipulating sex, consensus, distinctiveness, and role either as actor or observer found that raters are likely to project their own performance when they do not have consensus information. Fedor and Rowland (1989) proposed that if the raters perceive the ratee's behavior is deviating from how they (the rater) would perform in that situation, ratee performance is perceived negatively and low performance is attributed internally.

For raters with an internal locus of control, positive swells will be attributed internally resulting in higher attributions of judged ability and effort and higher performance ratings, whereas negative swells will be judged more harshly.

H11: Rater locus of control will moderate the relationship between tremors and (a) judged effort and (b) performance such that raters with an internal locus of control will provide higher ratings than raters with an external locus of control.

H12: Rater locus of control will moderate the relationship between positive swells and (a) judged ability, (b) judged effort, and (c) performance such that a *positive swell will be judged more favorably by raters with an internal locus of control than raters with an external locus of control.*

H13: Rater locus of control will moderate the relationship between negative swells and (a) judged ability, (b) judged effort, and (c) performance such that a negative swell will be judged more unfavorably by raters with an internal locus of control than raters with an external locus of control.

METHOD

Participants, Design, and Procedure

Participants were undergraduate students from a large university in the southwest. Professors of two classes from the Industrial Distribution department were contacted and asked to offer their students the opportunity to participate in the current study, with the option of awarding extra credit as determined by the instructors. Power analysis (p =.95) considering a small effect size of .1 (to detect even the smallest effect size) yielded a sample size requirement of 90 using G*Power (Faul, Erdfelder, Buchner & Lang, 2009). One hundred sixty-eight students participated in the study. Incomplete responses were discarded, reducing the sample to 148. The majority (78.4%) of the respondents were male (116) and the age of the respondents ranged from 18 to 36 years with an average of 21.68 years (SD = 2.91).

A 3 (mean level of performance: below average, above average, average) × 3 (swells: positive, negative, and zero) × 2 (tremors: low, high) × 2 (rater locus of control: internal, external) mixed experimental design was conducted. Mean, swells, and tremors were within-subjects factors, and rater locus of control was a between-subjects factor. Participants were asked to assume the role of a regional supervisor and rate the performance of 18 different junior salespersons. They were presented with 18 performance profiles that depict sales performance over the course of 26 weeks in a random order to avoid order effects. The performance profiles were similar to those used in Reb and Cropanzano's (2007) study with the same means and standard deviations. Gridlines were included in the profiles to make comparisons between profiles easier.

Instructions provided to the participants appear in Appendix A, and the 18 profiles appear in Appendix B.

Measures

Attributions. Attribution ratings were gathered for each of the 18 performance profiles. Single item measures adapted from Ronis, Hansen, and O'Leary (1983) and used in Reb and Greguras's (2010) study were used. To maximize the opportunity for variance in the judgments and ratings, consistent with Lee and Dalal (2011) an 11-point scale was used. Judged *ability* was evaluated with the following item: "Over the 26 weeks, this employee showed [strong/poor] ability." Judged *effort* was rated with the following item: "Over the 26 weeks, this employee invested [a lot of/little] effort." *Locus of causality* was assessed based on ratings of the following item: "Over the 26 weeks, this employee's performance was determined largely by factors [inside/outside] his/her control". Higher numbers on this scale were associated with more external attributions.

Performance. Raters evaluated performance on an 11-point scale (11 = very good) with the following item: "Over the 26 weeks, this employee had [very poor/very good] overall performance."

Locus of control. Rater locus of control was measured using a set of 20 bipolar items from the International Personality Item Pool (IPIP) ($\alpha = .86$) on a scale of 1-5, with one indicating external locus of control and five indicating internal locus of control (Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006).

Manipulations

Participants were told that the performance distributions are displayed relative to normative data gathered across numerous salespersons over multiple years, with the organization mean depicted as zero on the graph. Consistent with Reb and Cropanzano (2007), mean revenue contribution was (–$1,800; $0; or +$1,800), to represent below average, mean, or above average levels of performance. Standard deviation values of $200 and $600 were used to represent low and high levels of tremors. Individual data points for 26 weeks were simulated based on a low or high variation condition. Swells were created by adding or subtracting $2000 to any three consecutive weeks and consistent with Lee and Dalal's (2011) placement of outliers, were placed between weeks 16-21 to avoid recency and primacy effects. The presence of extremities near the midpoint causes a largely disproportionate impact (Gersick, 1988), hence the swells were positioned between weeks 16-21 to avoid this effect. Although such extremities are possible in the real world, for this study they were excluded.

RESULTS

A mixed measures analysis of variance on different dependent measures was conducted with mean level of performance (below average, above average, average), tremors (low and high), and extreme score swells (negative, positive and none) as within-subjects variables and locus of control (internal vs. external) as a between-subjects variable. Except for a few cases, the sphericity assumption was not violated. In cases where the sphericity assumption was violated, Hyunh-Feldt corrections were applied. Since the results from applying corrections were identical to those with sphericity assumed, results are reported with sphericity assumed. The means and standard deviations of the dependent measures are provided in Tables 3-6 and the analysis of variance of different dependent measures are provided in Tables 7-10.

For locus of control, along with the data from current study (n = 148, M = 3.67, SD = *.50*), additional data using the same scale from Deshpande, Payne, and Zoghi (2010) (n = 171, M = 4.42, SD = *.42*) and another sample (Lewis Goldberg, personal communication, March 23, 2015) (n = 725, M = 3.89, SD = *.42*) were used to determine the locus of control internal-external split. The grand mean across the three samples was calculated (M = 3.95), and participants scoring less than the mean were coded as externals, whereas those scoring more than mean were coded as internals.

Table 3

Means and Standard Deviations of Performance Ratings by Experimental Condition

Tremors	Swells	Mean Level of Performance: Below Average		Average		Above Average	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Low Tremors	Negative	2.59	1.46	4.82	1.53	7.71	1.54
	None	2.84	1.76	5.41	1.64	8.03	1.71
	Positive	3.29	1.46	5.98	1.48	8.42	1.45
High Tremors	Negative	2.41	1.33	4.46	1.61	7.78	1.63
	None	2.87	1.47	6.07	1.34	8.04	1.70
	Positive	3.28	1.66	6.49	1.53	8.84	1.64

Table 4

Means and Standard Deviations of Attributions of Ability by Experimental Condition

Tremors	Swells	Mean Level of Performance					
		Below Average		Average		Above Average	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Low Tremors	Negative	2.59	1.52	4.99	1.47	7.98	1.49
	None	2.96	1.79	5.54	1.64	8.19	1.58
	Positive	3.40	1.64	6.17	1.52	8.39	1.59
High Tremors	Negative	2.45	1.50	4.55	1.59	7.89	1.66
	None	2.89	1.47	6.21	1.23	8.19	1.57
	Positive	3.36	1.79	6.55	1.45	8.79	1.60

Table 5

Means and Standard Deviations of Attributions of Effort by Experimental Condition

		Mean Level of Performance					
Tremors	Swells	Below Average		Average		Above Average	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Low Tremors	Negative	2.82	1.65	4.85	1.51	7.57	1.68
	None	2.97	1.77	5.30	1.75	7.61	2.03
	Positive	3.54	1.72	5.99	1.59	8.18	1.64
High Tremors	Negative	2.88	1.89	4.70	1.75	7.80	1.66
	None	3.41	1.84	6.20	1.38	8.05	1.61
	Positive	3.61	1.92	6.45	1.52	8.66	1.66

Table 6

Means and Standard Deviations of Attributions of Locus of Causality by Experimental Condition (1 = Within the Employee's Control, 11 = Beyond the Employee's Control)

		Mean Level of Performance					
Tremors	Swells	Below Average		Average		Above Average	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Low Tremors	Negative	4.01	2.34	5.11	2.29	5.34	2.44
	None	3.65	2.27	4.77	2.19	4.84	2.71
	Positive	4.19	2.21	5.13	2.28	5.32	2.60
High Tremors	Negative	4.20	2.59	4.71	2.24	5.51	2.62
	None	4.01	2.33	5.12	2.28	5.14	2.64
	Positive	4.32	2.50	5.16	2.30	5.35	2.98

Table 7

Analysis of Variance Results for Performance Ratings

Source	*df*	*SS*	*MS*	*F*	η_p^2
Tremors	1	14	14	7.08**	0.046
Tremors * LoC	1	6	6	2.86	0.019
Error (Tremors)	146	295	2		
Swell	2	472	236	151.02**	0.508
Swell * LoC	2	4	2	1.15	0.008
Error (Swell)	292	456	2		
Mean Level	2	11,414	5,707	886.13**	0.859
Mean Level * LoC	2	4	2	0.27	0.002
Error (Mean Level)	292	1,881	6		
Tremors * Swell	2	29	15	12.83**	0.081
Tremors * Swell * LoC	2	7	3	2.98	0.020
Error (Tremors*Swell)	292	332	1		
Tremors * Mean Level	2	12	6	3.93*	0.026
Tremors * Mean Level * LoC	2	0	0	0.11	0.001
Error (Tremors*Mean Level)	292	448	2		
Swell * Mean Level	4	77	19	13.66**	0.086
Swell * Mean Level * LoC	4	3	1	0.49	0.003
Error (Swell*Mean Level)	584	823	1		
Tremors * Swell * Mean Level	4	25	6	5.34**	0.035
Tremors * Swell * Mean Level * LoC	4	10	3	2.18	0.015
Error (Tremors*Swell*Mean Level)	584	687	1		

Note. LoC = Rater locus of control. * $p < .05$. ** $p < .01$.

Table 8

Analysis of Variance Results for Attributions of Ability

Source	*df*	*SS*	*MS*	*F*	η_p^2
Tremors	1	6	6	2.80	0.019
Tremors * LoC	1	5	5	2.48	0.017
Error(Tremors)	146	320	2		
Swell	2	423	212	137.97**	0.486
Swell * LoC	2	4	2	1.31	0.009
Error(Swell)	292	448	2		
Mean Level	2	11,602	5,801	957.96**	0.868
Mean Level * LoC	2	7	4	0.60	0.004
Error(Mean Level)	292	1,768	6		
Tremors * Swell	2	32	16	13.31**	0.084
Tremors * Swell * LoC	2	13	6	5.41**	0.036
Error(Tremors*Swell)	292	348	1		
Tremors * Mean Level	2	9	5	3.34*	0.022
Tremors * Mean Level * LoC	2	1	0	0.27	0.002
Error(Tremors*Mean Level)	292	411	1		
Swell * Mean Level	4	85	21	13.94**	0.087
Swell * Mean Level * LoC	4	2	0	0.25	0.002
Error(Swell*Mean Level)	584	893	2		
Tremors * Swell * Mean Level	4	28	7	5.24**	0.035
Tremors * Swell * Mean Level * LoC	4	3	1	0.64	0.004
Error(Tremors*Swell*Mean Level)	584	774	1		

Note. LoC = Rater locus of control. * $p < .05$. ** $p < .01$.

Table 9

Analysis of Variance Results for Attributions of Effort

Source	*df*	*SS*	*MS*	*F*	η_p^2
Tremors	1	70	70	16.18**	0.100
Tremors * LoC	1	2	2	0.36	0.002
Error(Tremors)	146	631	4		
Swell	2	375	188	87.85**	0.376
Swell * LoC	2	6	3	1.40	0.009
Error(Swell)	292	624	2		
Mean Level	2	9,474	4,737	642.22**	0.815
Mean Level * LoC	2	8	4	0.52	0.004
Error(Mean Level)	292	2,154	7		
Tremors * Swell	2	31	15	10.40**	0.066
Tremors * Swell * LoC	2	13	7	4.51*	0.030
Error(Tremors*Swell)	292	433	1		
Tremors * Mean Level	2	8	4	2.25	0.015
Tremors * Mean Level * LoC	2	2	1	0.65	0.004
Error(Tremors*Mean Level)	292	509	2		
Swell * Mean Level	4	77	19	11.48**	0.073
Swell * Mean Level * LoC	4	6	1	0.87	0.006
Error(Swell*Mean Level)	584	976	2		
Tremors * Swell * Mean Level	4	17	4	2.79	0.019
Tremors * Swell * Mean Level * LoC	4	1	0	0.20	0.001
Error(Tremors*Swell*Mean Level)	584	897	2		

Note. LoC = Rater locus of control. * $p < .05$. ** $p < .01$.

Table 10

Analysis of Variance Results for Attributions of Locus of Causality

Source	*df*	*SS*	*MS*	*F*	η_p^2
Tremors	1	13	13	1.91	0.013
Tremors * LoC	1	3	3	0.38	0.003
Error(Tremors)	146	1,015	7		
Swell	2	64	32	7.79**	0.051
Swell * LoC	2	47	24	5.77**	0.038
Error(Swell)	292	1,196	4		
Mean Level	2	560	280	30.53**	0.173
Mean Level * LoC	2	55	27	2.98	0.020
Error(Mean Level)	292	2,678	9		
Tremors * Swell	2	15	8	3.06*	0.021
Tremors * Swell * LoC	2	0	0	0.10	0.001
Error(Tremors*Swell)	292	722	2		
Tremors * Mean Level	2	6	3	1.14	0.008
Tremors * Mean Level * LoC	2	0	0	0.04	0.000
Error(Tremors*Mean Level)	292	818	3		
Swell * Mean Level	4	20	5	2.02	0.014
Swell * Mean Level * LoC	4	4	1	0.39	0.003
Error(Swell*Mean Level)	584	1,412	2		
Tremors * Swell * Mean Level	4	12	3	1.25	0.008
Tremors * Swell * Mean Level * LoC	4	6	2	0.69	0.005
Error(Tremors*Swell*Mean Level)	584	1,368	2		

Note. LoC = Rater locus of control. * $p < .05$. ** $p < .01$.

Main Effects

Hypothesis 1 stated that there will be a main effect for mean level of performance on attributions of locus of causality such that lower levels of performance are more likely to be attributed internally than externally. As predicted, mean level of performance had a significant main effect on locus of causality, F (2, 292) = 30.53, $p < .001$, $\eta_p^2 = .17$. Profiles with below- average mean level of performance (M = 4.03) were significantly more likely to be attributed internally than average mean level of performance (M = 4.93) and above-average mean level of performance (M = 5.12). Thus, Hypothesis 1 was supported.

Hypotheses 2 and 3 stated that there will be a positive main effect for mean level of performance on judged ability and on judged effort. As predicted, mean level of performance had a significant main effect on ability, F (2, 292) = 957.96, $p < .001$, $\eta_p^2 = .87$, and effort F (2, 292) = 642.22, $p < .001$, $\eta_p^2 = .81$. Profiles with above average mean level of performance were rated the highest on judged ability (M = 8.22), followed by average mean level of performance (M = 5.62), and below average mean level of performance (M = 2.91). Similarly, for judged effort, profiles with above average mean level of performance were rated the highest (M = 7.97), followed by average mean level of performance (M = 5.53), and then below average mean level of performance (M = 3.17). Thus, Hypotheses 2 and 3 were supported.

Hypothesis 4 stated that there will be a positive main effect for the mean level of performance on performance ratings. As predicted, mean level of performance had a significant main effect on performance ratings, F (2, 292) = 886.13, $p < .001$, $\eta_p^2 = .86$.

For performance ratings, profiles with above average mean level of performance were rated the highest (*M* = 7.11), followed by average mean level of performance (*M* = 5.49), and then below average mean level of performance (*M* = 2.85). Thus, Hypothesis 4 was also supported.

Research Question 1 focused on the effect of tremors on attributions of locus of causality, whereas Research Question 2 focused on the effect of tremors on ratings of performance. There was no significant effect for tremors on attributions of locus of causality F (1, 146) = 1.91, $p = .17$, $\eta_p^2 = .01$. A high level of tremors led to significantly higher ratings of performance (*M* = 5.41) than a low level of tremors (*M* = 5.56), F (1,146) = 7.08, $p < .01$, $\eta_p^2 = .05$.

Hypothesis 5 stated that there will be a positive main effect for tremors on judged effort. As predicted, a high level of tremors led to higher attributions of effort (*M* = 5.39) than a low level of tremors (*M* =5.73), F (1,146) = 16.14, $p < .001$, $\eta_p^2 = .10$, thus, Hypothesis 5 was also supported.

Hypothesis 6 stated that there will be a main effect for swells on attributions of locus of causality such that compared to positive swells, negative swells are more likely to be attributed internally. There was a significant main effect of swells on attributions of locus of causality, such that profiles with a positive swell (*M* = 4.83) and negative swell (*M* = 4.79) were significantly more likely to be attributed externally than profiles without a swell (*M* = 4.47) F (2,292) = 7.79, $p < .001$, $\eta_p^2 = .05$. However, profiles with a negative swell were not more likely to be attributed internally compared to those with a positive swell. Thus, Hypothesis 6 was not supported. Moreover, attributions of locus of causality for both negative and positive swells were opposite of what was predicted such that negative swells were more likely to be attributed externally and positive swells were more likely to be attributed internally.

Hypothesis 7 stated that the presence of a positive swell will lead to higher ratings of (a) performance, (b) judged ability, and (c) judged effort compared to profiles without swells, whereas Hypothesis 8 stated lower ratings for negative swells compared to profiles without swells. As predicted, profiles with positive swells received higher ratings for performance ($M = 6.01$), $F(2,292) = 151.02$, $p < .001$, $\eta_p^2 = .51$, judged ability ($M = 6.07$) $F(2,292) = 137.97$, $p < .001$, $\eta_p^2 = .49$, and judged effort ($M = 6.04$), $F(2,292) = 87.84$, $p < .001$, $\eta_p^2 = .38$, whereas those with negative swells received significantly lower ratings ($M = 4.94$; $M = 5.06$; $M = 4.79$, respectively) as compared to profiles without any swells. Thus, Hypotheses 7 and 8 were supported.

Performance Variation Interactions

Hypothesis 9 stated that there will be an interaction between tremors and mean level of performance such that a high level of tremors will have a positive effect on ratings when performance is above average, whereas a high level of tremors will have a negative effect on ratings when performance is below average. As predicted and depicted in Figure 1, there was a significant interaction between level of performance and tremors, $F(2,294) = 3.93$, $p < .05$, $\eta_p^2 = .03$, such that for profiles with below average performance, profiles with a high level of tremors ($M = 2.83$) were rated significantly lower than profiles with a low level of tremors ($M = 2.86$). For profiles with above average performance, performance was rated higher for profiles with a high level of tremors (M = 8.21) than for profiles with low level of tremors (M = 8.02). Thus, Hypothesis 9 was supported.

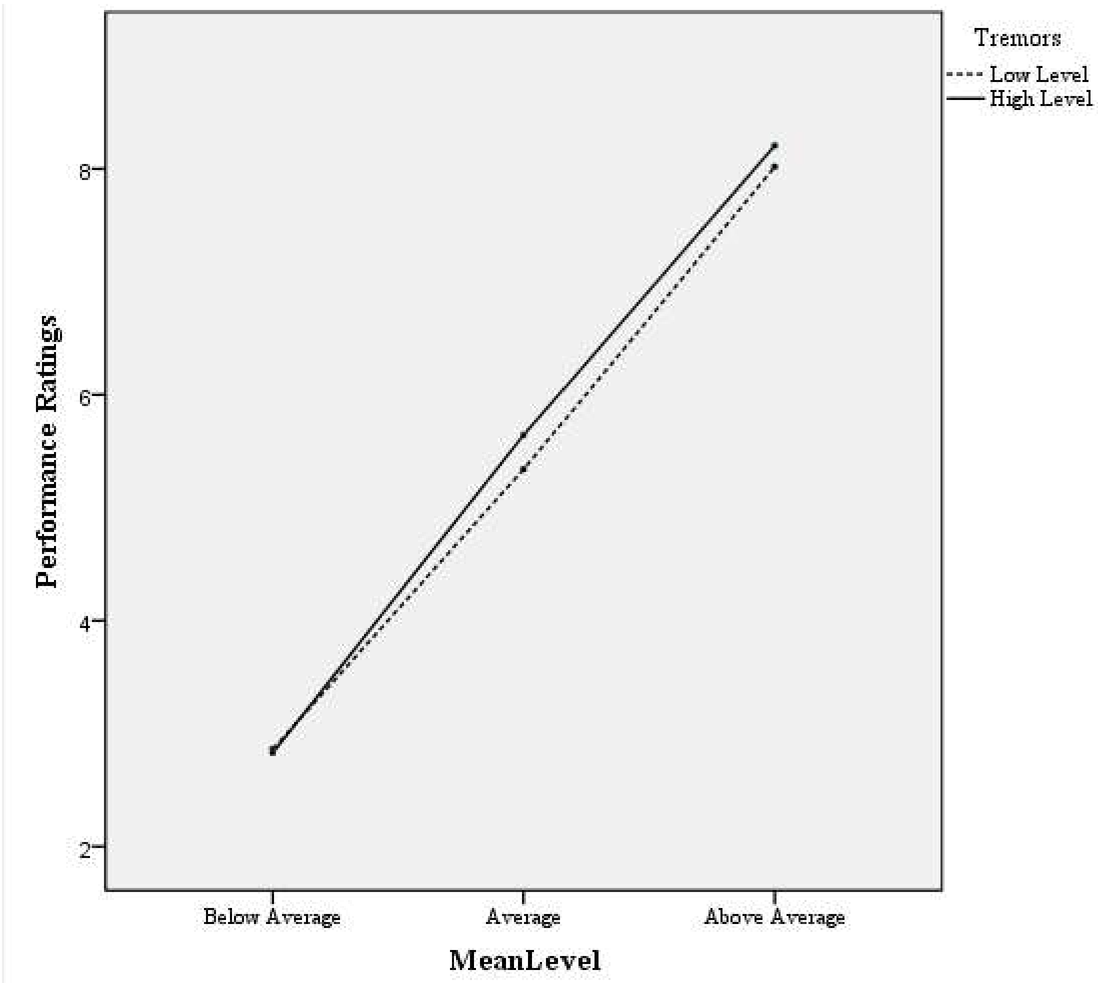

Figure 1. Interaction between tremors and mean level of performance on performance ratings.

Hypothesis 10 stated that there will be an interaction between tremors and swells on performance ratings such that high level of tremors will have a positive effect on ratings when combined with (a) positive swells and a negative effect on ratings when combined with (b) negative swells compared to profiles without swells. As predicted and depicted in Figure 2, there was a significant interaction of swells and tremors, F (2,294) = 12.83, $p < .001$, $\eta_p^2 = .08$, such that profiles with a high level of tremors and positive swells (M = 4.87) were rated significantly lower than those with low level of tremors (M = 5.01). Profiles with positive swells and a high level of tremors (M = 6.20) were rated significantly higher than profiles with a low level of tremors (M = 5.83). Thus, Hypothesis 10 was supported.

Rater Locus of Control and Causal Attributions

Hypothesis 11 stated that rater locus of control will moderate the relationship between tremors and (a) judged effort and (b) performance such that raters with an internal locus of control will provide higher ratings than raters with an external locus of control. There was not a significant interaction between rater locus of control and performance variability for attributions of effort, F (1,146) = .36, *ns* or performance ratings, F (1,146) = 2.86, *ns*. Thus, Hypothesis 11 was not supported.

Hypothesis 12 stated that rater locus of control will moderate the relationship between positive swells and (a) judged ability, (b) judged effort, and (c) performance such that a positive swell will be judged more favorably by raters with an internal locus of control than raters with an external locus of control. Hypothesis 13 stated that rater locus of control will moderate the relationship between negative swells and (a) judged ability, (b) judged effort, and (c) performance such that a negative swell will be judged more negatively by raters with an internal locus of control than raters with an external locus of control. There were no significant interactions between rater locus of control and swells for attributions of ability F (2,294) = 1.31, p =.27, η_p^2 = .01, effort F (2,294) = 1.40, *ns* or performance ratings F (2,294) = 1.154, *ns*. Thus, Hypotheses 12 and 13 were not supported.

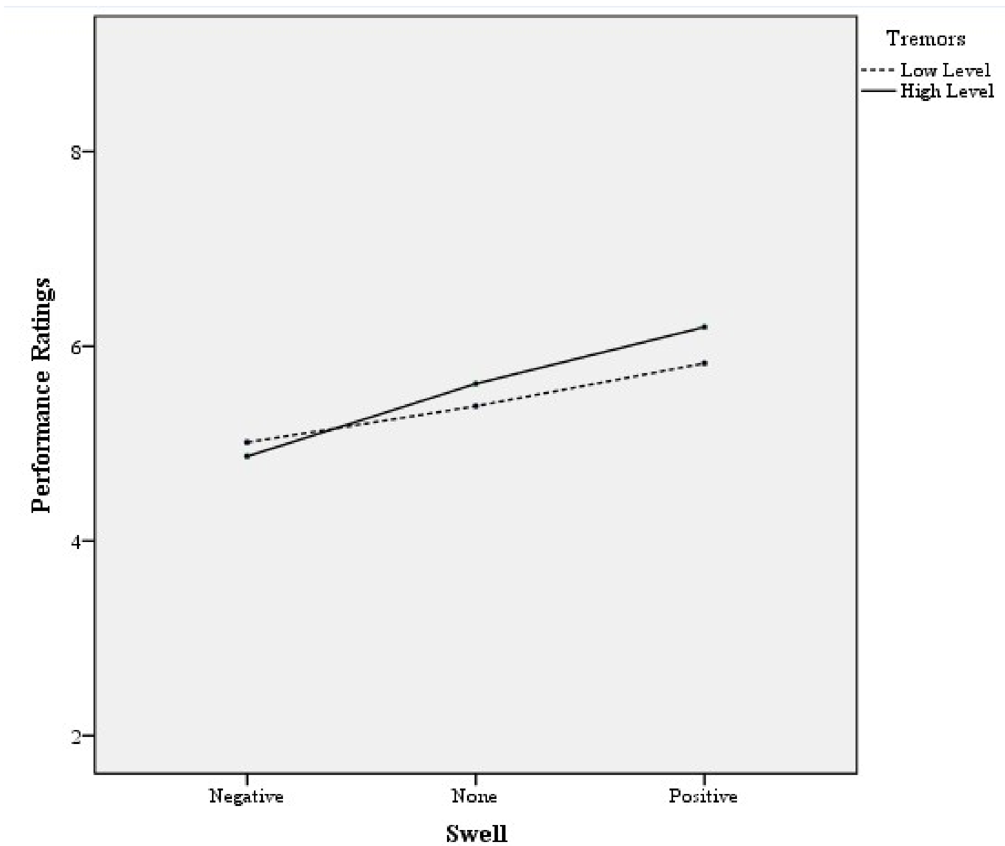

Figure 2. Interaction between tremors and swells on performance ratings.

Though not hypothesized, as depicted in Figure 3, there was a significant interaction between rater locus of control and swells on attributions of locus of causality,

F (2,294) = 5.77, $p < .01$, $\eta_p^2 = .04$ such that raters with internal locus of control were more likely to rate profiles with a negative swell (M = 4.68), no swell (M = 4.03) and a positive swell (M = 4.53) internally than raters with external locus of control (M = 4.89; M = 4.91; M = 5.13, respectively).

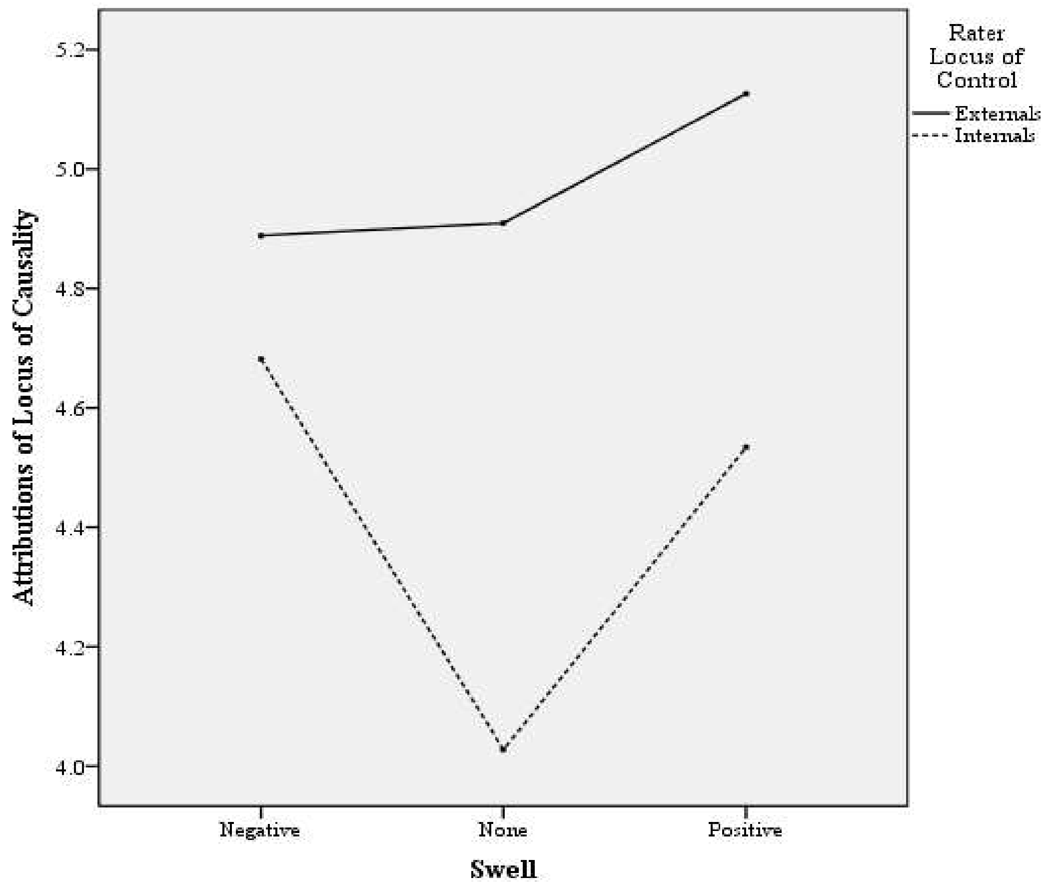

Figure 3. Interaction between rater locus of control and swells on rater attributions of locus of causality. Smaller numbers on the attributions of causality scale (1-11) indicate causality due to factors within the employee’s control, whereas larger numbers indicate causality due to external factors beyond the employee’s control.

DISCUSSION

The current study was conducted with the intent of contributing to the understanding of the extent to which performance variability affects supervisor ratings of performance as well as supervisors' attributions about the ratee's ability, effort, and locus of causality. Whereas earlier experiments examined variability either as tremors (using standard deviations) or trends (DeNisi & Stevens, 1981; Lee & Dalal, 2011, Reb & Cropanzano, 2007; Reb & Greguras, 2010; Scott & Hamner, 1975), to the best of my knowledge, this was the first experiment to manipulate two types of variability: short-term variability or tremors (using standard deviation), and longer-term variability or swells (using extreme score clusters). As predicted, results supported the hypotheses that swells have a significant impact on ratings of performance and attributions of ability and effort, with positive swells having a positive impact and negative swells having a negative impact.

As expected, mean level of performance had the strongest effect on performance ratings with a partial eta squared (η_p^2) of .85. A high mean level of performance was associated with higher attributions of ability, effort, and performance. Swells had the next strongest effect on performance ratings with a η_p^2 of .51. Positive swells led to higher attributions of ability and effort, whereas negative swells led to lower attributions of ability and effort. The effect sizes for both the mean level of performance as well as swells can be interpreted as large (Cohen, 1988).

Furthermore, contrary to Reb and Greguras (2010) who found a negative relationship between tremors and ratings of performance, and Reb and Cropanzano

(2007) who found no relationship between tremors and performance ratings, in the current study, tremors had a positive effect on performance ratings. In this study, the tremors were manipulated using the same values as Reb and Cropanzano ($200 as standard deviation for small variation, $ 600 for large variation) which were doubled in Reb and Greguras's study. It is possible that some level of variability may be acceptable and even rewarded when it is in a positive direction because it shows potential for improvements, whereas a large level of variability may be penalized. It is also possible that since Reb and Cropanzano also manipulated trends in their study, trends were much more salient and overrode any effects that the tremors might have had.

There was not a significant difference in rater's attributions of locus of causality for positive vs. negative swells. Although, it was predicted that profiles with swells were more likely to be attributed internally than profiles without swells, the results did not support this prediction. This might be due to the fact that no information was given regarding the causes of positive and negative swells. Since raters were more likely to make more external attributions for profiles with swells than profiles without swells, it appears that the raters may have assumed external factors such as luck or market conditions contributed to those swells.

Contrary to the results of Lee and Dalal (2011) who found a significant effect for negative extreme scores on performance ratings but not for positive extreme scores, both positive and negative swells had a significant effect on performance ratings. This difference may be due to cultural differences in the samples of raters as Lee and Dalal examined a sample in a (eastern) South Korean University setting, whereas the current study examined a sample in an American University (western) setting. Lee and Dalal pointed out that in eastern cultures, positive performance extremity can be interpreted as "showing off and as a threat to harmony" (p. 113). A sample of raters that varies considerably on the individualism dimension of culture (Hofstede, 2001) would need to be tested to further explore this speculation as North America scores a very high 91 points as compared to South Korea which scores 18 on the individualism scale (Hofstede, Hofstede, & Minkov, 2010)

The difference in results between the current study and Lee and Dalal's (2011) study may also be due to differences in the manipulation of the outliers. In the current study, swells were operationalized as three consecutive outliers, whereas in Lee and Dalal's study, outliers were operationalized as one or two data points that were not consecutive. As anticipated, it appears that three outliers in a row is more salient to the rater and warrants reacting to when rating the profile of performance, regardless of whether it is positive or negative.

As predicted, there was a significant interaction between short-term variability (tremors) and mean level of performance and tremors and swells on performance ratings. Compared to profiles with a low level of tremors, profiles with a high level of tremors and below average level of performance were given lower ratings, whereas profiles with an average and above average level performance were given higher ratings. Similarly, compared to profiles with a low level of tremors, profiles with a high level of tremors and a negative swell received lower ratings, whereas profiles with a positive swell received higher ratings. Thus, it appears that high variability was penalized when combined with below average performance or a negative swell, whereas it was rewarded when combined with above average performance or a positive swell. A three-way interaction between mean level of performance, tremors, and swells illuminates these results further. A combination of below average level of performance, a high level of tremors and a negative swell resulted in the lowest ratings of performance, whereas the combination of above average performance, a high level of tremors and a positive swell led to the highest ratings of performance. In short short-term variability is tolerated more so when combined with above average performance and/or longer term positive variability.

Also contrary to the predictions, rater locus of control did not moderate relationships between tremors and performance ratings or attributions of ability or effort, nor did rater locus of control moderate relationships between swells and performance ratings or attributions of ability or effort. However, rater locus of control did moderate the relationship between swells and attributions of locus of causality, such that raters with an internal locus control were significantly more likely to attribute performance internally as compared to raters with an external locus of control. This is consistent with Reid and Ware (1973) who found that locus of control was transferrable to ratings of others. It is unlikely that ratees will have much control over who rates them and the rater's locus of control, but raters can be made aware of their own locus of control and how that might sway the attributions that they make. More research is needed to reveal when differences in rater locus of control and attributions result in differences in ratings.

Theoretical Implications

In this study, raters did not rate a single episode of performance but rather a performance profile which was a temporal map of performance over time. The temporal mapping of performance provided raters with the shape of the performance profile and information about performance variability (tremors, and swells) that had a significant effect on performance ratings as well as attributions of ability and effort. Considering the well-documented impact of trend on performance ratings from previous studies (e.g., Reb & Cropanzano, 2007), and the large effect size associated with swells in this study, it can be concluded that intraindividual performance variability does influence performance ratings.

As Reb and Cropanzano (2007) pointed out, these results do not reduce the importance of the mean level of performance that accounted for the largest amount of variance in performance ratings in their study as well as the current study. Instead the interaction between the mean and variability reveals that the mean is not the only variable considered by raters when making ratings and that the mean can have a slightly stronger effect in the presence of variability.

The current study is the first study to manipulate swells. Consistent with Dalal et al.'s (2014) typology of performance variability, the current study's results also support that raters rate each of the multiple forms of performance variability (tremors and swells) differently as the effect of swells on performance ratings is stronger than the effect of tremors. Further, whereas tremors positively affected attributions of effort, positive/negative swells affected attributions of both ability and effort. Raters attributed higher levels of effort to performance profiles with a large amount of tremors. They also attributed higher/lower levels of ability and effort to performance profiles with positive/negative swells. Swells had a significant effect on rater attributions of ability as well as attributions of effort such that raters attributed positive swells to higher levels of ability and effort and negative swells to lower levels of ability and effort.

The current study illustrates that performance variability is not a unidimensional construct. The influence of tremors on ratings is different than the influence of swells and the influence of swells depends on where it appears in the performance distribution relative to the mean (i.e., positive or negative). In order to truly understand the influence of performance variability on performance ratings, each dimension of performance variability that Dalal et al. identified (linearity, duration, permanence, and cycle) needs to be examined individually.

With regard to Gestalt characteristics, below average performance was more likely to be attributed internally than average or above average performance. Thus, raters were more likely to attribute performance internally than externally when performance was poor. This is consistent with the fundamental attribution error (Ross, 1977) and the results of Carson et al. (1991).

Practical Implications

Results from Reb and Cropanzano (2007), Reb and Greguras (2010) and the current study are completely different from each other with regard to the effect of tremors on performance ratings. Reb and Cropanzano did not find a significant relationship between tremors and performance ratings; Reb and Greguras found a negative relationship; whereas the current study found significant positive relationship. It may be that tremors have a curvilinear relationship with performance ratings rather than a linear one. This relationship may take on a bell-shaped curve or inverted U-shape such that initially performance variability is positively related to performance ratings, but upon reaching an upper limit of variability that is tolerated, raters may give lower ratings for increasing performance variability (e.g., a very high level of tremors).

To further complicate things, individual raters may tolerate different levels of variability. In other words, each rater may have a different threshold when it comes to variability. Correspondingly, each rater may choose to ignore, reward, or punish a different level of tremors within a performance profile. Future research is needed to see if raters vary in their thresholds for variability and what predicts this variability.

Since variability affects performance ratings, it would not be appropriate to compare two profiles just on the basis of mean level of performance. Research to date suggests a profile with a low mean level of performance and high level of tremors, or upward trend or presence of a positive outlier is likely to be rated higher than a profile with a high mean and negative outlier. A rater may be faced with multiple combinations of such scenarios based on different levels of means and different types of variability. Organizations, therefore, may need to educate raters on all of these possibilities and possibly design performance appraisal systems that capture not only the mean level of performance but also variability. Organizations should provide clear guidelines as to the levels of tremors (e.g., standard deviation of sales), swells, and trends tolerated or encouraged/punished and how they should be rated.

The characteristics of the temporal performance profile provide a wealth of information for various managerial needs (Reb & Cropanzano, 2007). Different types of variability in the performance profile can be analyzed for decisions regarding bonuses, developmental feedback, and identifying problems that may not be possible if only the mean level of performance were considered.

Limitations and Future Research Directions

The current study utilized an experimental design allowing for the manipulation of the variables of interest at specific levels predicted to have an influence on performance ratings as well as combinations of variables to ensure a balanced experimental design. Nevertheless, the current study has some limitations that warrant acknowledgement. This study was conducted using student raters rating hypothetical employees performing only one task and such simulated performance profiles may or may not match the performance profiles of the actual salespersons. In order to keep the amount of information presented to the raters manageable, no information was provided to the raters regarding the causes for the swells or tremors. In contrast, when supervisors evaluate employees on the job, they are likely to have at least some information about what may be contributing to performance variability and if they don't have sufficient information, they could potentially seek out more information. Field studies using raters with supervisory experience rating current employees would provide a more ecologically valid rating experience, but may not present all possible combinations of variables of interest. Further, other variables that may influence performance ratings like the tenure of rater in the supervisory role, the quality of rater-ratee relationship, the amount of time the ratee has worked for the supervisor, and the organizational culture could also be measured.

Future studies should include multiple levels of variability to determine if the relationship of tremors with performance ratings is linear or curvilinear and at what threshold variability is perceived negatively. Although the current study had many significant findings, the current study used the same 11-point Likert scale as Lee and Dalal (2011). For greater differentiation across profile ratings and attributions, the use of a continuous scale with a slider may enhance such differentiation even further, especially when teasing out differences between experimental conditions to determine the relationship between tremors and performance ratings across multiple levels of tremors. Similarly, interactions between tremors and mean as well as tremors and swells should also be examined.

The current study provided performance profiles to raters in graphical format. Some research has shown a significant interaction between mean and display type (graphical vs. tabular), such that compared to a graphical display, an above average level of performance was rated more positively and below average performance was rated more negatively in a tabular condition (Reb & Cropanzano, 2007). On the other hand, for the significant interactions between trend and display type, the graphical display received more positive ratings as compared to the tabular type, except for the deteriorating trend where both of them had equal ratings (Reb & Cropanzano, 2007). Given the difference in results for the interaction of display type with mean and trend, comparisons of graphical versus tabular format should also be done to see the differences in outcomes for interactions of display type with swells and also to test how it affects the replication of the Reb and Cropanzano's study in the presence of swells.

In the current study, rater locus of control moderated the relationship between swells and rater attribution of locus of causality, such that raters with an internal locus of control were more likely to attribute the swells internally than raters with an external locus of control. However, rater locus of control did not moderate the relationship between swells and attributions of effort or performance ratings. This may be due to the fact that no reasons or causes for performance variability were provided. Furthermore, a single item was used to measure the influence of internal versus external causes. Further studies should provide competing reasons (internal and external causes) for the causes of swells as well as measure attributions of each cause independently in order to further examine the impact of rater locus of control.

Finally, most of the job performance profile studies done are in the areas of sales. Similar studies should be carried out in other fields with high variability in performance such as sports to examine the effects of the three types of variability (tremors, swells, and trends) in other industries.

In conclusion, this study set out to examine the effect of two forms of performance variability (short-term tremors and long-term swells) and rater locus of control, on supervisory ratings of performance and attributions of ability, effort and locus of causality. Results revealed that along with the mean, swells had significant effect on performance ratings as well as rater attributions of ability, effort and locus of causality. Although tremors had a significant main effect on performance ratings as well as attributions of ability, no significant relationship was found between tremors and attributions of locus of causality. Rater locus of control moderated the relationship between swells and attributions of locus of causality confirming that locus of control was transferrable to ratings of others. These findings suggest that performance variability is a complex multidimensional construct and raters take note of at least two of these dimensions: duration and permanence. Considering the results of the current and previous performance variability studies, organizations may need to provide raters with explicit guidance on how to evaluate different forms of variability.

REFERENCES

Anderson, C. A. (1983). Motivational and performance deficits in interpersonal settings: The effect of attributional style. *Journal of Personality and Social Psychology*, *45(5)*, 1136-1147. http://dx.doi.org/10.1037/0022-3514.45.5.1136

Ariely, D., & Carmon, Z. (2000). Gestalt characteristics of experiences: The defining features of summarized events. *Journal of Behavioral Decision Making, 13*, 191–201. http://dx.doi.org/10.1002/(SICI)1099-0771(200004/06)13:2%3C191:: AID-BDM330%3E3.0.CO;2-A

Ariely, D., & Carmon, Z. (2003). Summary assessment of experiences: The whole is more than the sum of its parts. In G. Loewenstein, D. Read, & R. Baumeister (Eds.), *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 323–349). New York: Russell Sage Foundation.

Ashkanasy, N. M. (1989). Causal attribution and supervisors' responses to subordinate performance: The Green and Mitchell model revisited. *Journal of Applied Social Psychology, 19*, 309-330.

Bar-tal, D., Goldberg, M., & Knaani, A. (1984). Causes of success and failure and their dimensions as a function of SES and gender: A phenomenological analysis. *British Journal of Educational Psychology, 54,* 51-61. http://dx.doi.org/10.1111/ j.2044-8279.1984.tb00844.x

Barnes, C. M., & Morgeson, F. P. (2007). Typical performance, maximal performance, and performance variability: Expanding our understanding of how organizations value performance. *Human Performance, 20*, 259–274. http://dx.doi.org/10.1080/

08959280701333289

Beal, D. J., Weiss, H. M., Barros, E., & MacDaniel, S. M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology, 90,* 1054–1068. http://dx.doi.org/10.1037/ 0021-9010.90.6.1054

Burger, J. M., Cooper, H. M., & Good, T. L. (1982). Teacher attributions of student performance: Effects of outcome. *Personality & Social Psychology Bulletin, 4*, 685-690. http://dx.doi.org/10.1177/0146167282084013

Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: A consideration of alternative perspectives*. Cincinnati, OH: South-Western Publishing Co.

Carson, K. P., Cardy, R. L., & Dobbins, G. H. (1991). Performance appraisal as effective management or deadly management disease: Two initial empirical investigations. *Group and Organization Studies, 16*, 143-159. http://dx.doi.org/10.1177/ 105960119101600203

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology, 74*, 130-135. http://dx.doi.org/10.1037/0021-9010.74.1.130

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Connell III, H. (1977). What to do when the selling slump hits. *Training & Development Journal*, *31*(11), 14.

Cooper, H. M. & Burger, J. M. (1980). How teachers explain students. *American Educational Research Journal, 17(1)*, 95-109. http://dx.doi.org/10.3102/

00028312017001095

Dalal, R. S., Bhave, D. P., & Fiset, J. (2014). Within-person variability in job performance: A theoretical review and research agenda. *Journal of Management*, *40*, 1396-1436. http://dx.doi.org/ 10.1177/0149206314532691

Deadrick, D. L., Bennett, N., & Russell, C. J. (1997). Using hierarchical linear modeling to examine dynamic performance criteria over time. *Journal of Management, 23,* 745–757. http://dx.doi.org/10.1177/014920639702300603

Deadrick, D. L., & Madigan, R. M. (1990). Dynamic criteria revisited: A longitudinal study of performance stability and predictive validity. *Personnel Psychology, 43*, 717–744. http://dx.doi.org/10.1111/j.1744-6570.1990.tb00680.x

Deming, W. E. (1986). *Out of the crisis*. Cambridge, Mass: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance, 6*, 360-395. http://dx.doi.org/10.1016/0030-5073(84)90029-1

DeNisi, A. S., & Stevens, G. E. (1981). Profiles of performance, performance evaluations, and personnel decisions. *Academy of Management Journal*, *24*, 592–602. http://dx.doi.org/ 10.2307/255577

Deshpande, S. S., Payne, S. C., & Zoghi B. (2010, April). *Narrow personality predictors of technical sales performance*. Poster paper presented at the 25th annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Elig, T. W., & Frieze, I. H. (1979). Measuring causal attibutions for success and failure. *Journal of Personality and Social Psychology*, *37(4)*, 621-634. http://dx.doi.org/10.1037/0022-3514.37.4.621

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. http://dx.doi.org/10.3758/BRM.41.4.1149

Fedor, D. B., & Rowland, K. M. (1989). Investigating supervisor attributions of subordinate performance. *Journal of Management, 15,* 405–416. http://dx.doi.org/10.1177/ 014920638901500304

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology, 66*, 127-148. http://dx.doi.org/ 10.1037/0021-9010.66.2.127

Fisher, C. D., & Noble, C. S. (2004). A within-person examination of correlates of performance and emotions while working. *Human Performance, 17*, 145–168. http://dx.doi.org/ 10.1207/s15327043hup1702_2

Frieze, I., & Weiner, B. (1971). Cue utilization and attributional judgments of success and failure. *Journal of Personality, 39*, 591-605. http://dx.doi.org/10.1111/ j.1467-6494.1971.tb00065.x

Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal*, *31*, 9–41. http://dx.doi.org/ 10.2307/256496

Gilbert, D. T., and P. S. Malone. 1995. The correspondence bias. *Psychological Bulletin, 117,* 21–38. http://dx.doi.org/10.1037/0033-2909.117.1.21

Gilmor, T. M., & Minton, H. L. (1974). Internal versus external attribution of task performance as a function of locus of control, initial confidence, and success-failure outcome. *Journal of Personality, 42*, 159-174. http://dx.doi.org/10.1111/ j.1467-6494.1974.tb00563.x

Gilmor, T. M., & Reid, D. W. (1979). Locus of control and causal attribution for positive and negative outcomes on university examinations. *Journal of Research in Personality, 13(2)*, 154-160. http://dx.doi.org/10.1016/0092-6566(79)90027-8

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, England: Cambridge University Press.

Ghiselli, E. E., & Haire, M. (1960). The validation of selection tests in light of the dynamic nature of criteria. *Personnel Psychology, 13*, 225-231. http://dx.doi.org/10.1111/ j.1744-6570.1960.tb01352.x

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84-96. http:// dx.doi.org/10.1016/j.jrp.2005.08.007

Green, S. G., & Mitchell, T. R. (1979). Attributional processes in leader– member interactions. *Organizational Behavior and Human Performance, 23*, 429–458. http://dx.doi.org/ 10.1016/0030-5073(79)90008-4

Hansen, R. D., & Lowe, C. A. (1976). Distinctiveness and consensus: The influence of behavioral information on actors' and observers' attributions. *Journal of Personality and Social Psychology, 34*, 425-433. http://dx.doi.org/10.1037/0022-3514.34.3.425

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self–supervisor, self–peer, and peer–supervisor ratings. *Personnel Psychology, 41,* 43–62. http://dx.doi.org/ 10.1111/ j.1744-6570.1988.tb00631.x

Harvey, P., Madison, K., Martinko, M.J., Crook, T.R., & Crook, T. (2014). Attribution theory in the organizational sciences: The road traveled and the path ahead. *Academy of Management Perspectives, 28,* 128-146. http://dx.doi.org/ 10.5465/amp.2012.0175

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons.

Hoffman, D. A., Jacobs, R., & Gerras, S. J. (1992). Mapping individual performance over time. *Journal of Applied Psychology, 77,* 185–195. http://dx.doi.org/ 10.1037/0021-9010.77.2.185

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2^{nd} ed.), Thousand Oaks CA: Sage Publications.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind* (Rev. 3rd ed.). New York: McGraw-Hill.

Hulin, C. L., Henry, R., & Noon, S. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relations. *Psychological Bulletin, 107,* 328–340. http://dx.doi.org/10.1037/0033-2909.107.3.328

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance, *Psychological Bulletin, 96*, 72-98. http://dx.doi.org/10.1037/0033-2909.96.1.72

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus., In Staw, B. M., and Cummings, L. L. (eds.), *Research in organizational behavior*, (pp.141-197). JAI Press, Greenwich, CT.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79-94). Morristown, NJ: General Learning Press.

Kahneman, D. (2000). Evaluation by moments: Past and future. In D. Kahneman & A. Tversky (Eds.), *Choices, values and frames* (pp. 693–708). New York: Cambridge University Press and the Russell Sage Foundation.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Kane, J. S. (1984). Performance distribution assessment: A new breed of appraisal methodology. In H. J. Bernardin & R. W. Beatty (Eds.), *Performance appraisal: Assessing human behavior at work* (pp. 325-341). Boston: Kent.

Kane, J. S. (1986). Performance distribution assessment. In R. A. Berk (Ed.), *Performance assessment* (pp. 237–273). Baltimore: Johns Hopkins University Press.

Kelley, H. H. (1972). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (eds.), *Attribution: Perceiving the causes of behavior* (pp. 1-26). Morristown, NJ: General Learning Press.

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist, 28*, 107-128. http://dx.doi.org/10.1037/h0034225

Knowlton, W. A., & Mitchell, T. R. (1980). Effects of causal attributions on a supervisor's evaluation of subordinate performance. *Journal of Applied Psychology, 65*, 459–466. http://dx.doi.org/10.1037/0021-9010.65.4.459

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt Brace.

Lee, H. & Dalal, R. S. (2011). The effects of performance extremities on ratings of dynamic performance. *Human Performance, 24*, 99-118. http://dx.doi.org/ 10.1080/ 08959285.2011.554467

Lefcourt, H. M. (1976). *Locus of control*. New York: Lawrence Erlbaum Associates.

Lefcourt, H. M., Hogg, E., Struthers, S., & Holmes C. (1975). Causal attributions as a function of locus of control, initial confidence, and performance outcomes. *Journal of Personality and Social Psychology, 2*, 391-397. http://dx.doi.org/ 10.1037/h0077090

Lumsden, J. (1977). Personal reliability. *Applied Psychological measurement, 1*, 477-482. http://dx.doi.org/10.1177/014662167700100403

Martinko, M. J. (Ed.). (1995). *Attribution theory: An organizational perspective*. Delray

Beach, FL: St. Lucie Press.

Martinko, M. J., Douglas, S. C., & Harvey, P. (2006). Attribution theory in industrial and organizational psychology: A review. In Hodgkinson, G. P. & Ford, J. K. (Eds.), *International review of industrial and organizational psychology* (Vol. 21, pp. 127–187). Chichester, UK: Wiley. http://dx.doi.org/10.1002/9780470696378.ch4

Mitchell, T. R., & Wood, R. E. (1980). Supervisor's responses to subordinate poor performance: A test of an attributional model. *Organizational Behavior and Human Performance, 25*, 123-138. http://dx.doi.org/10.1016/0030-5073(80)90029-X

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.

Nesselroade, J. R. (2001). Intraindividual variability in development within and between individuals. *European Psychologist, 6*, 187-193. http://dx.doi.org/10.1027/1016-9040.6.3.187

Reb, J., & Cropanzano, R. (2007). Evaluating dynamic performance: The influence of salient Gestalt characteristics on performance ratings. *Journal of Applied Psychology*, *92*, 490–499. http://dx.doi.org/10.1037/0021-9010.92.2.490

Reb, J., & Greguras, G. (2008). Dynamic performance and the performance-performance rating relation. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 194–196. http://dx.doi.org/10.1111/j.1754-9434.2008.00038.x

Reb, J., & Greguras, G. (2010). Understanding performance ratings: Dynamic performance, attributions, and rating purpose. *Journal of Applied Psychology*, *95*, 213–220. http://dx.doi.org/10.1037/a0017237

Reid, D. W., & Ware, E. E. (1973). Multidimensionality of internal- external control: Implications for past and future research. *Canadian Journal of Behavioral Science*, *5*, 264-271. http://dx.doi.org/10.1037/h0082351

Ronis, D. L., Hansen, R. D., & O'Leary, V. E. (1983). Understanding the meaning of achievement attributions: A test of derived locus and stability scores. *Journal of Personality and Social psychology, 44(4)*, 702-711. http://dx.doi.org/10.1037/ 0022-3514.44.4.702

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 174-221). New York: Academic Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*. (Whole No. 609). http://dx.doi.org/10.1037/h0092976

Russell, D., McAuley, E., & Tarico, V. (1987). Measuring causal attributions for success and failure: A comparison of methodologies for assessing causal dimensions. *Journal of Personality and Social Psychology, 52*, 1248-1257. http://dx.doi.org/ 10.1037/ 0022-3514.52.6.1248

Salancik, G.R. & Pfeffer, J. (1978). A social information processing approach to job attitudes and task design, *Administrative Science Quarterly 23*, 224-253.

http://dx.doi.org/ 10.2307/2392563

Scott, W. E., & Hamner, W. C. (1975). The influence of variations in performance profiles on the performance evaluation process: An examination of the validity of the criterion. *Organizational Behavior and Human Performance*, *14*, 360–370. http://dx.doi.org/ 10.1016/0030-5073(75)90036-7

Stewart, G. L., & Nandkeolyar, A. K. (2006). Adaptation and intraindividual variation in sales outcomes: Exploring the interactive effects of personality and environmental opportunity. *Personnel Psychology, 59*, 307–332. http://dx.doi.org/10.1111/j.1744-6570.2006.00038.x

Thoresen, C. J., Bradley, J. C., Bliese, P. D., & Thoresen, J. D. (2004). The Big Five personality traits and individual job performance growth trajectories in maintenance and individual stages. *Journal of Applied Psychology, 89*, 835–853. http://dx.doi.org/10.1037/ 0021-9010.89.5.835

U.S. Department of Labor, Bureau of Labor Statistics. (2014, March 1).*Occupational outlook handbook, Wholesale and Manufacturing Sales Representatives.* Retrieved from http://www.bls.gov/ooh/sales/ wholesale-and-manufacturing-sales-representatives.htm

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology, 87(2),* 345-354. http://dx.doi.org/10.1037/0021-9010.87.2.345

Weiner, B. (1974). Achievement motivation as conceptualized by an attribution theorist. In B. Weiner (Ed.), *Achievement motivation and attribution theory*. Morristown, N.J.: General Learning Press.

Weiner, B. (1985). An attribution theory of achievement motivation and emotion. *Psychological Review, 92,* 548–573. http://dx.doi.org/10.1037/ 0033-295X.92.4.548

Weiner, B., Frieze, I., Kulka, L., Reed, L., Rest, S., & Rosenbaum, R. M. (1971). *Perceiving the causes of success and failure.* In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (eds.), *Attribution: Perceiving the causes of behavior* (pp. 95-120), Morristown, NJ: General Learning Press.

Williams, J. (2014). *Graphical method on performance appraisals.* Retrieved from http://smallbusiness.chron.com/graphical-method-performance-appraisals-1193.html

Willson, V. L. & Palmer, D. J. (1983). Latent partition analysis of attributions for actual achievement. *American Educational Research Journal, 20,* 581-589. http://dx.doi.org/ 10.3102/00028312020004581

Wotruba, T. R., & Schoel, D. J. (1983). Evaluation of salesforce contest performance. *Journal of Personal Selling & Sales Management*, *3*, 1-10.

APPENDIX A

INSTRUCTIONS

Purpose:

The purpose of this study is to examine how supervisors rate the performance of employees. As you know, performance ratings are very important in determining the course of an individual's career. Thus, it is important for us to know how such ratings are made.

Information About Yourself

Please complete the questionnaire about yourself. This questionnaire contains some questions about you and your viewpoint about certain aspects that will help us understand your decision making.

Your Role:

In this present study, we would like you to play the role of a Regional Supervisor. You are in charge of a firm that supplies wholesale appliances to retail outlets. Under your supervision are 18 junior-level sales personnel. Your task is to review their performance over the past 26 weeks and to give them their semi-annual performance evaluation. These performance appraisals are used for personnel record keeping and to document your judgment of their overall performance over the pay period in question.

Your Information

You will base your judgments on data for the past 26 weeks. In other words, for each of the 18 salespersons you will see their performance over 26 weeks depicted in graphs. These weekly performance data for each salesperson show how much money they contributed to the company in dollar amounts. Specifically, the number for each week conveys how much sales revenue that person brought in relative to a long-term company average as measured over several years and many, many salespersons.

For example, in the example below, the salesperson generated revenue of about $200 *more* than the long-term company average in Week 1, and $200 *less* than the long-term company average in Week 2.

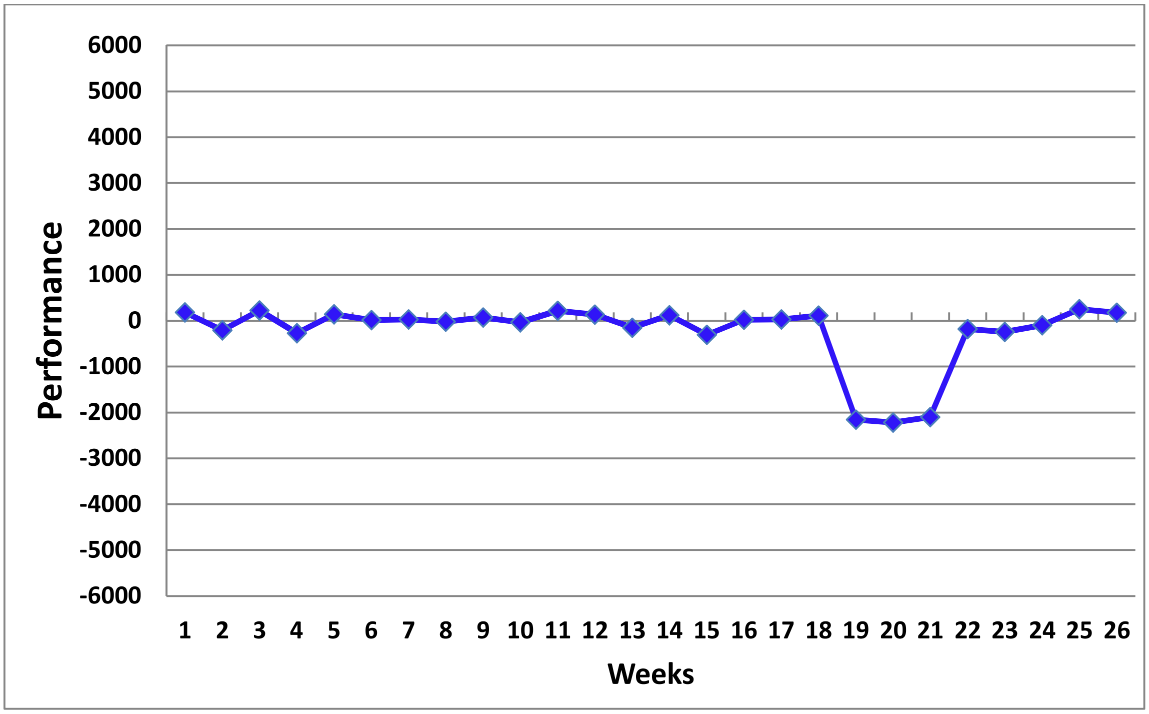

Figure A-1. Performance Profile Example

You will see one graph for each salesperson. In the graph, the weeks will be on the x-axis (labeled "week") and the revenue contribution will be on the y-axis (labeled "performance").

<u>Making Your Evaluations</u>

You have all the available information on the 18 employees you need. To complete your job, please enter your evaluations directly below each of the performance profiles. You will evaluate each employee on the same criteria.

APPENDIX B

Demographics:

Full Name: Class:

Race: White/Black/Hispanic/Asian/unspecified

Gender: Male/female Age:

Work Experience:

Please provide the number of times you have evaluated the following on your job

Self ___________

Peer _____________

Subordinate ___________

Please provide the number of times you have evaluated the following on your academic school?

Self ___________

Peer _____________

Subordinate ___________

Performance Profiles

Observe the following 18 performance profiles carefully and rate each of the junior salespersons (each profile).

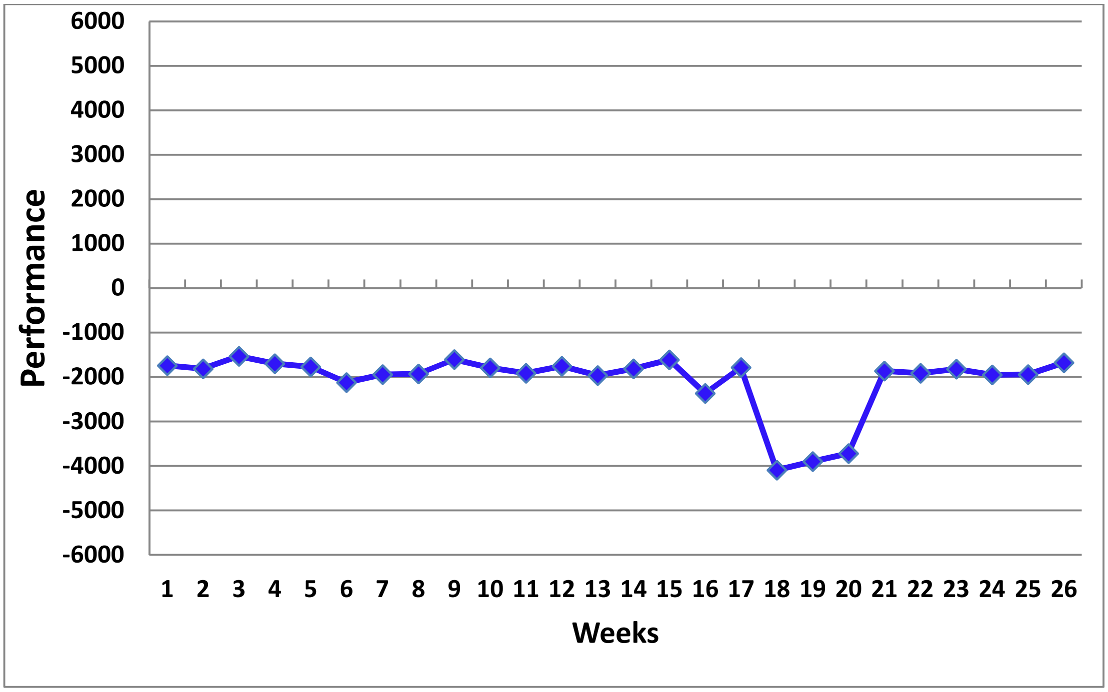

Figure A-2. Below average, low tremors, negative swell

Judged ability: "Over the past 26 weeks, this employee showed [strong/poor] ability"
1-extremely poor ability, 11-extremely strong ability

Judged effort: "Over the past 26 weeks, this employee invested [a lot of/little] effort"
1-extremely little effort, 11 –extremely high effort

Judged locus of causality: "Over the past 26 weeks, this employee's performance was determined largely by factors [inside/outside] his/her control".

1- Largely by factors within employee's control, 11-largely by factors beyond employee's control

Past Performance: "Over the past 26 weeks, this employee had [very poor/very good] overall performance."
1 – extremely poor performance, 11- extremely good performance

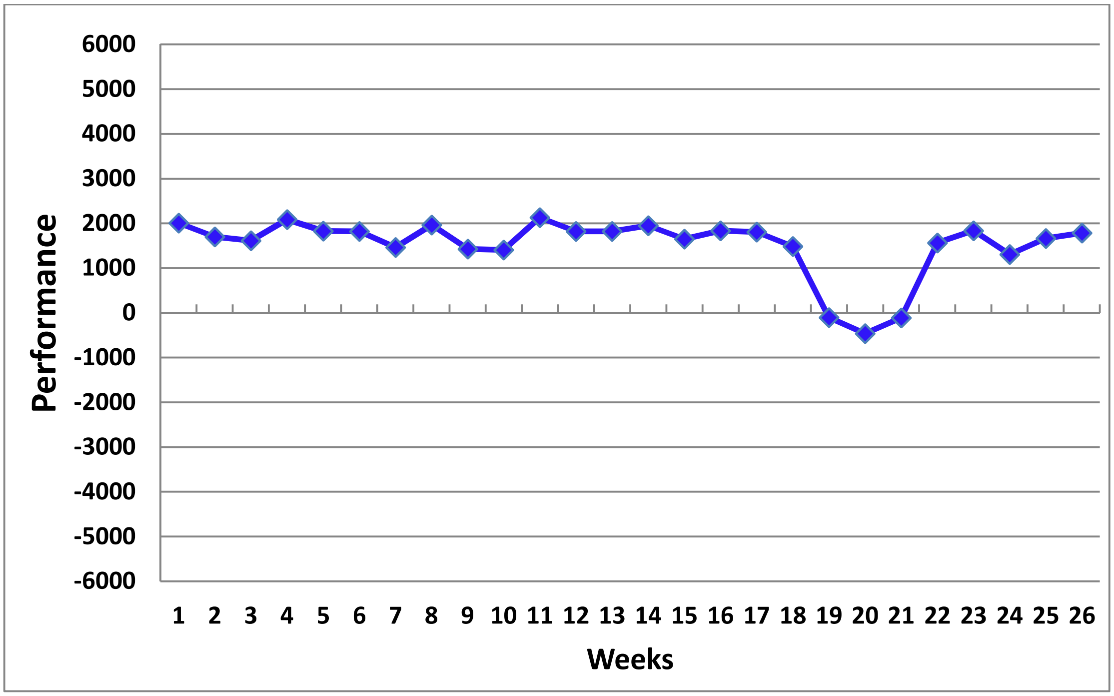

Figure A-3. Above average, low tremors, negative swell

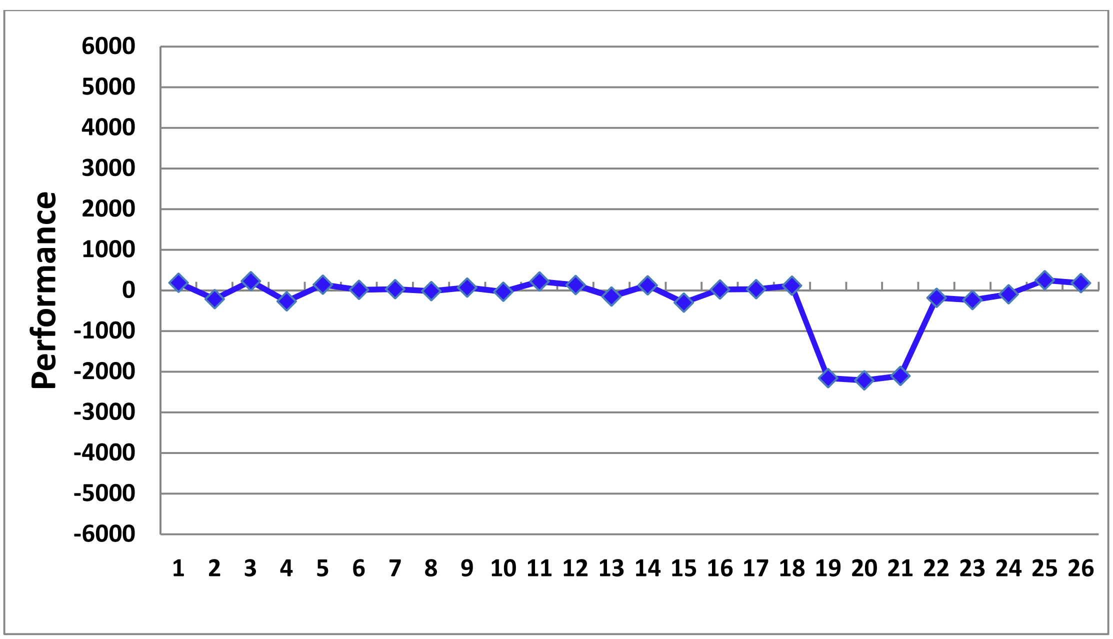

Figure A-4. Average, low tremors, negative swell

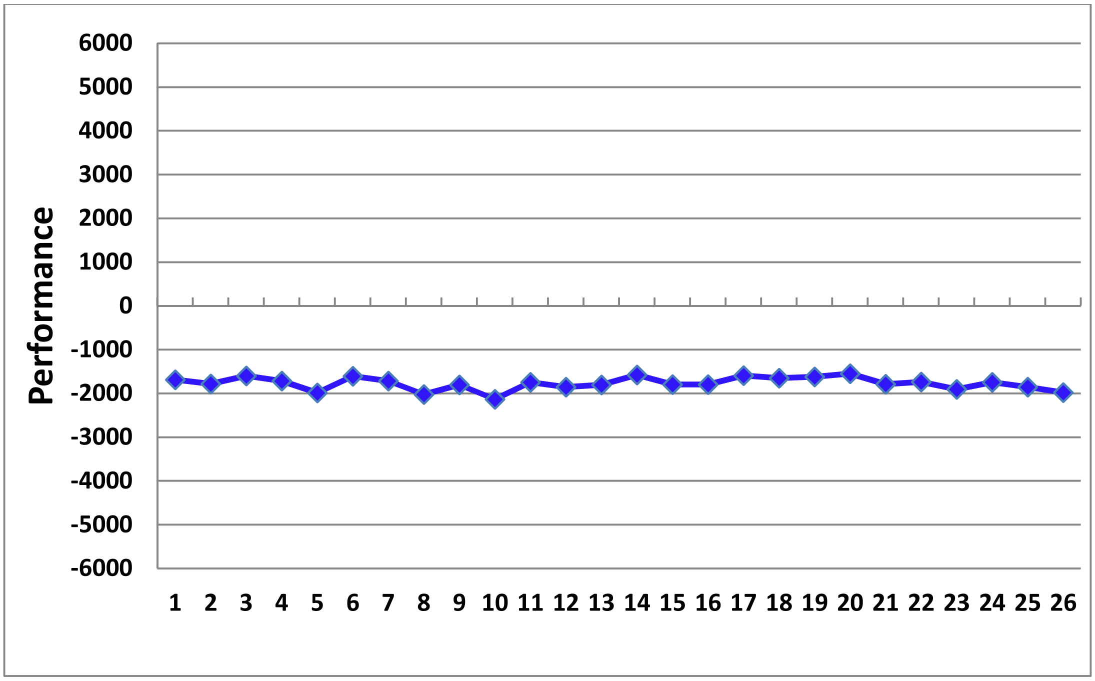

Figure A-5. Below average, low tremors, no swell

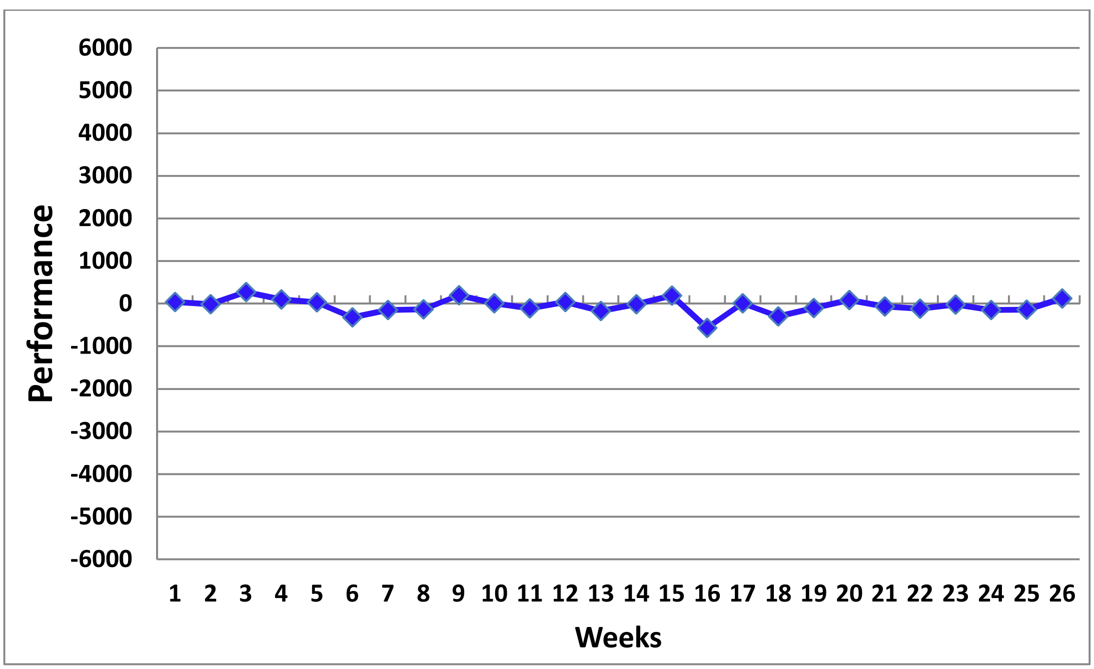

Figure A-6. Average, low tremors, no swell

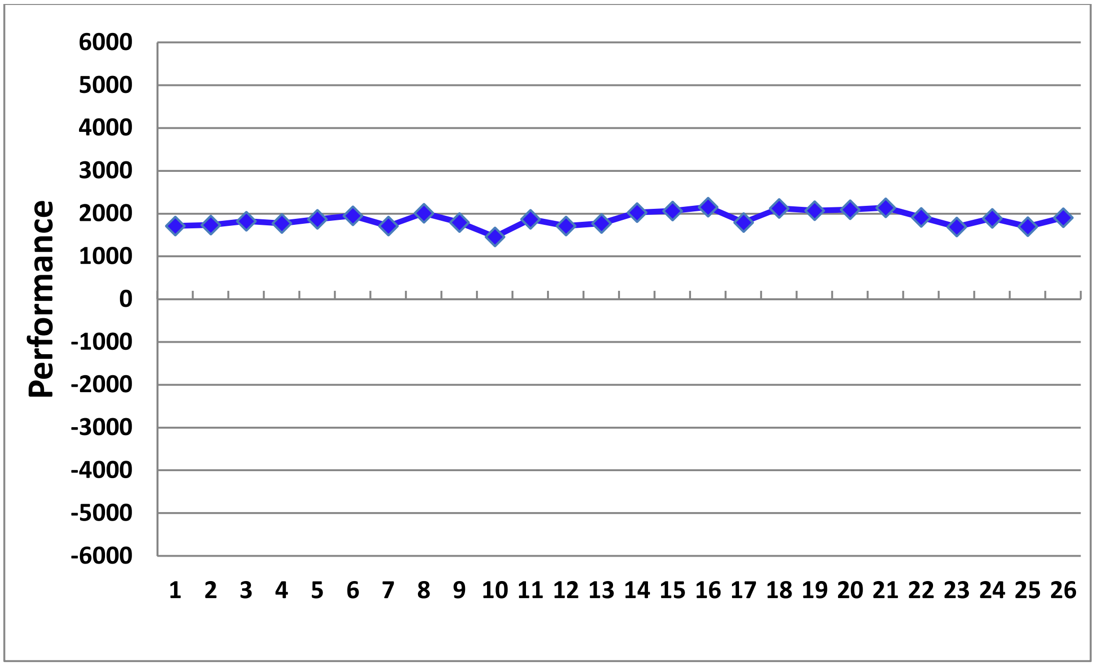

Figure A-7. Above average, low tremors, no swell

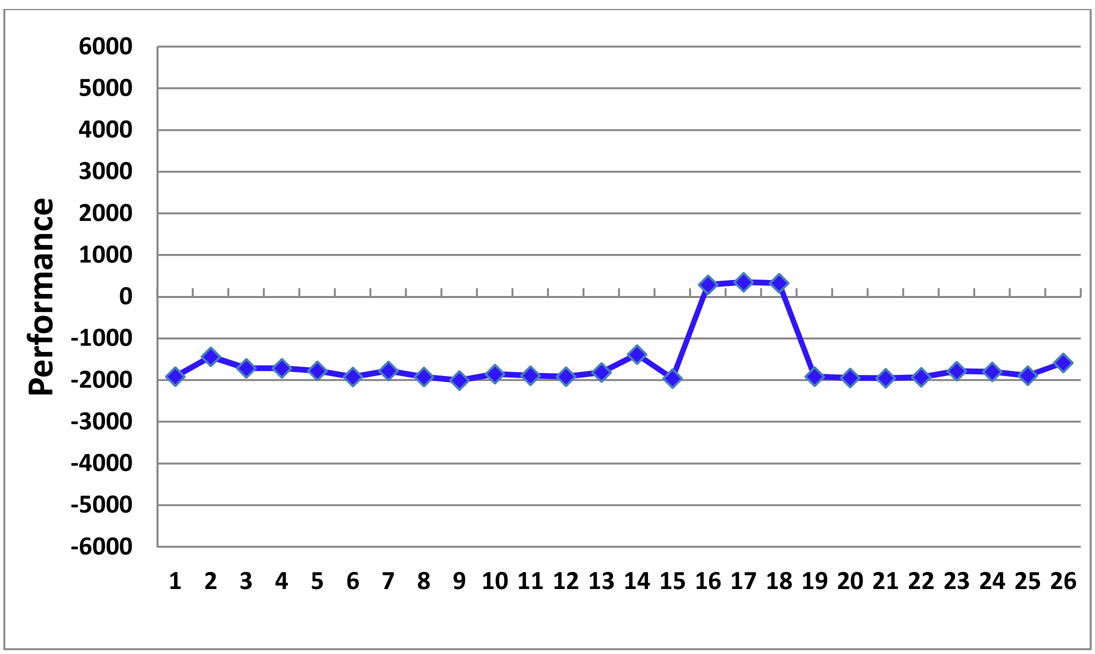

Figure A-8. Below average, low tremors, positive swell

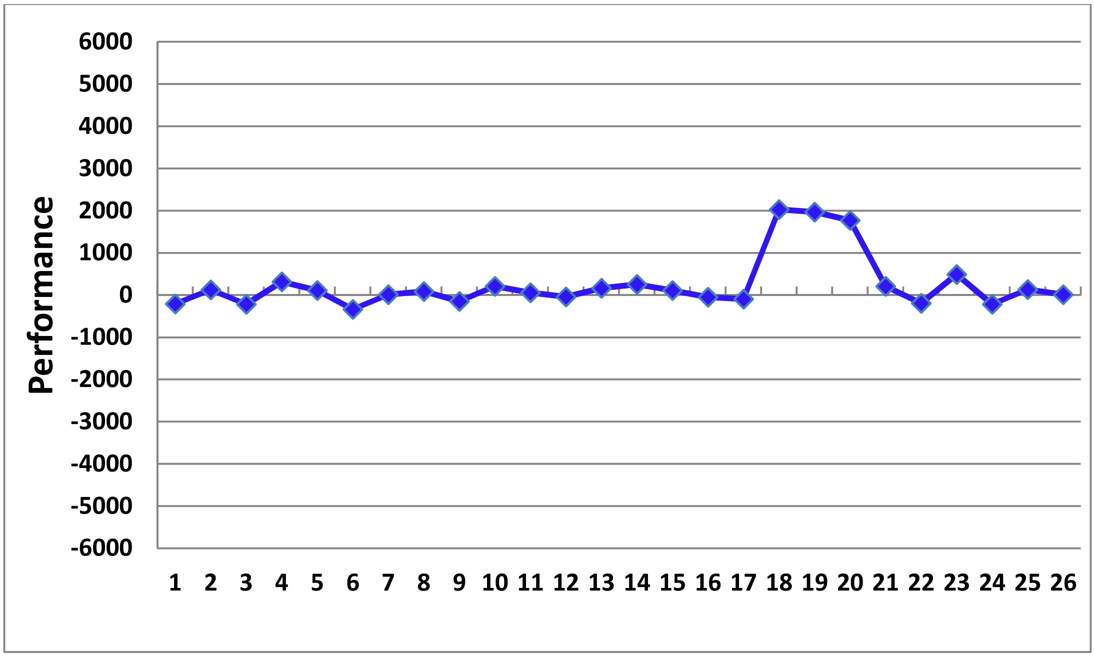

Figure A-9. Average, low tremors, positive swell

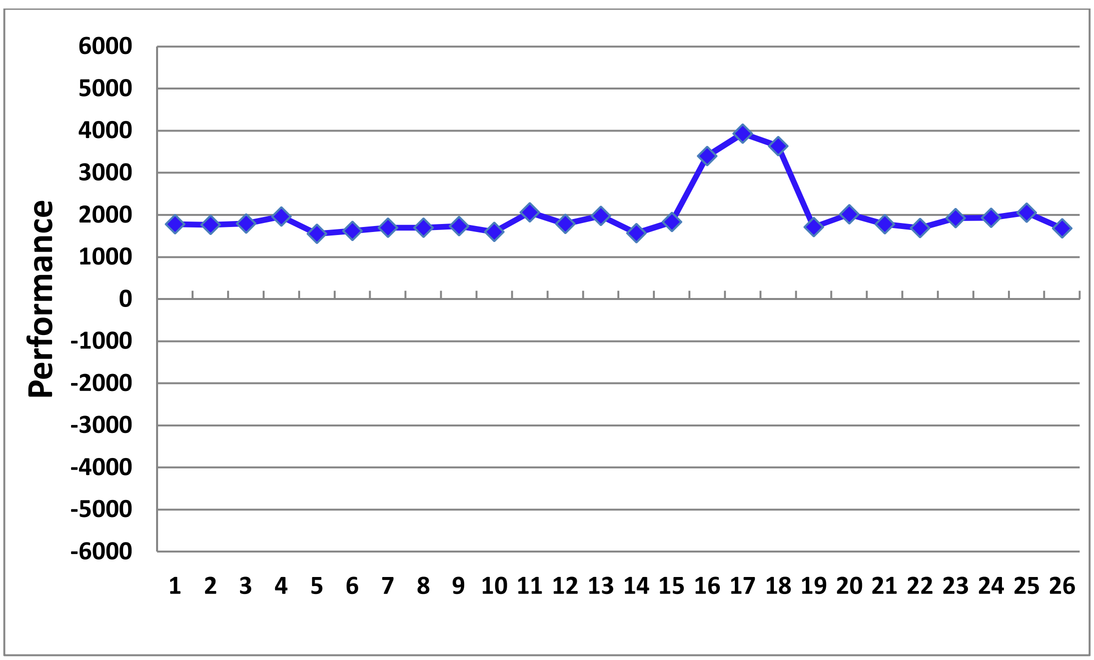

Figure A-10. Above average, low tremors, positive swell

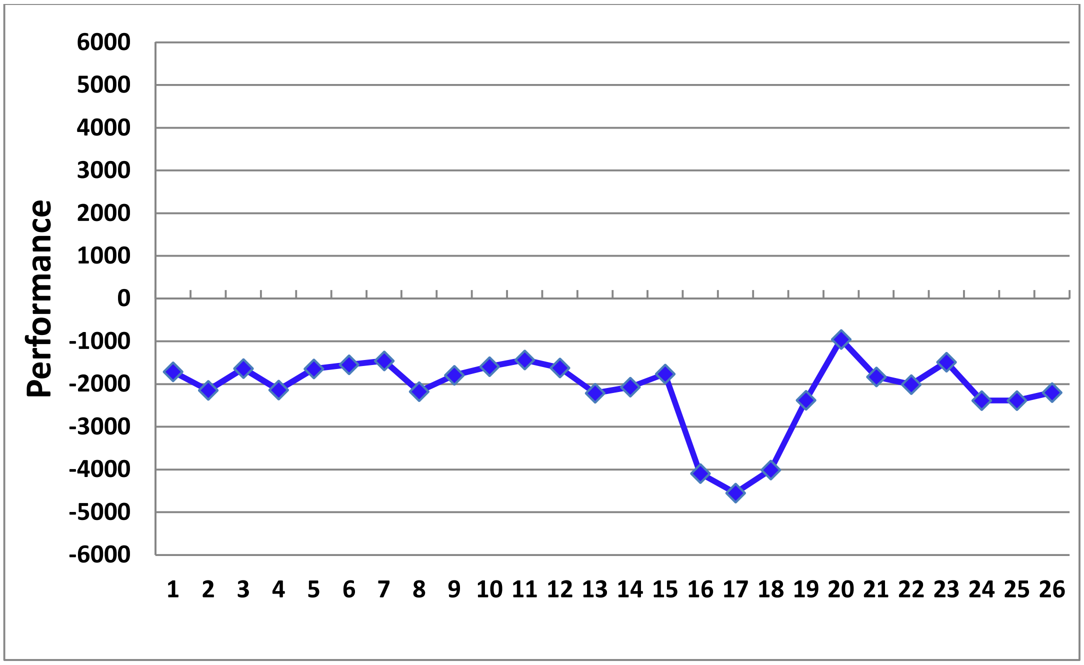

Figure A-11. Below average, high tremors, negative swell

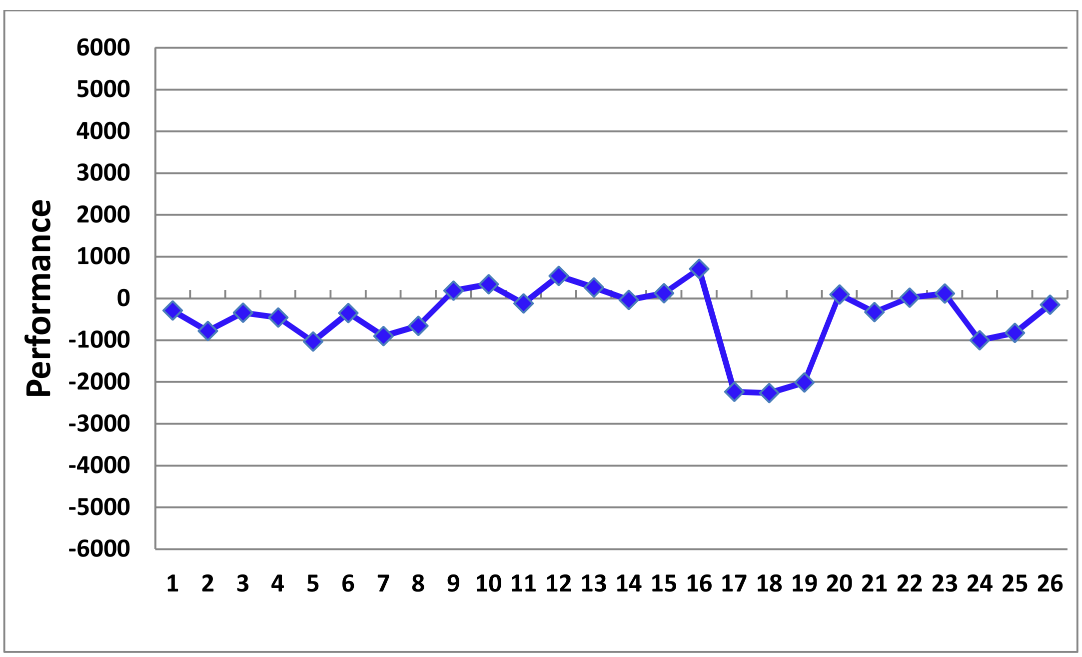

Figure A-12. Average, high tremors, negative swell

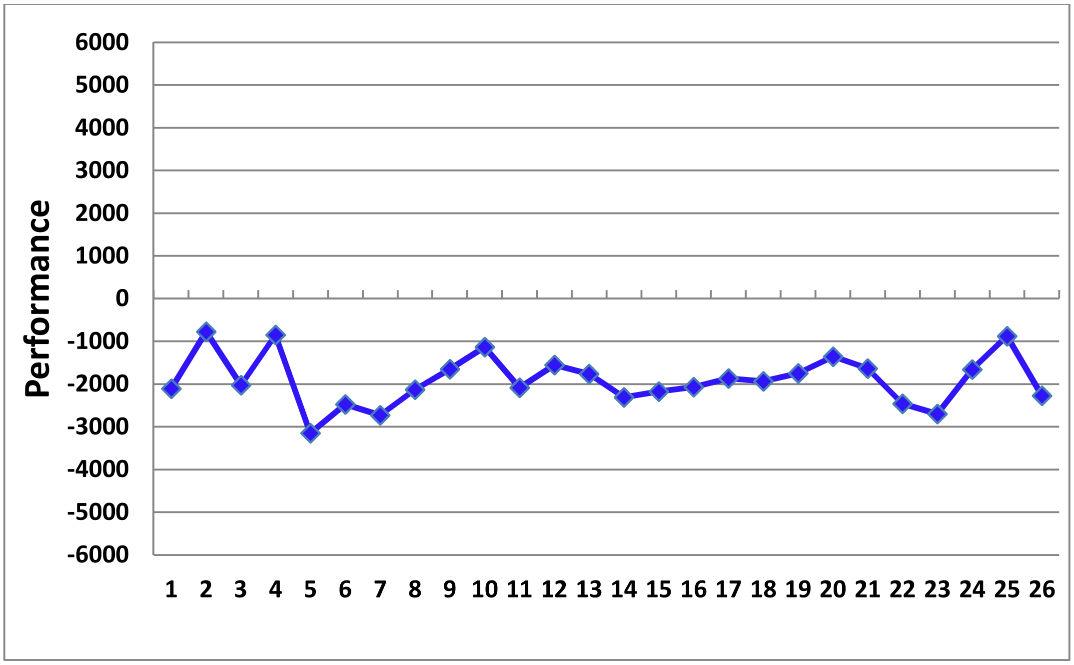

Figure A-13. Below average, high tremors, no swell

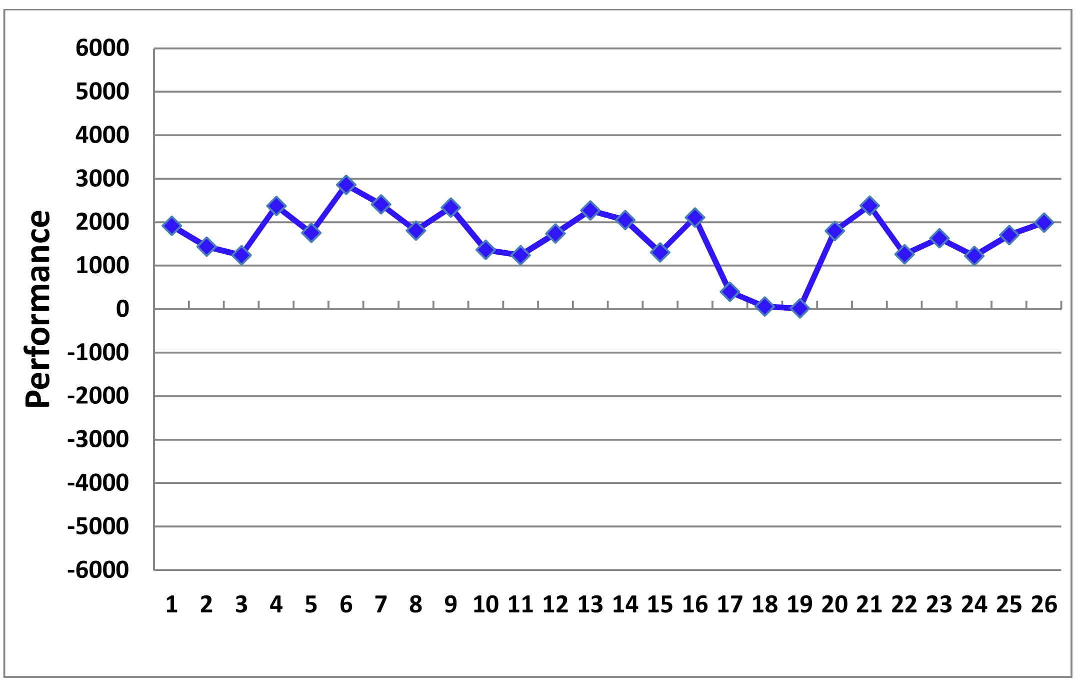

Figure A-14. Above average, high tremors, negative swell

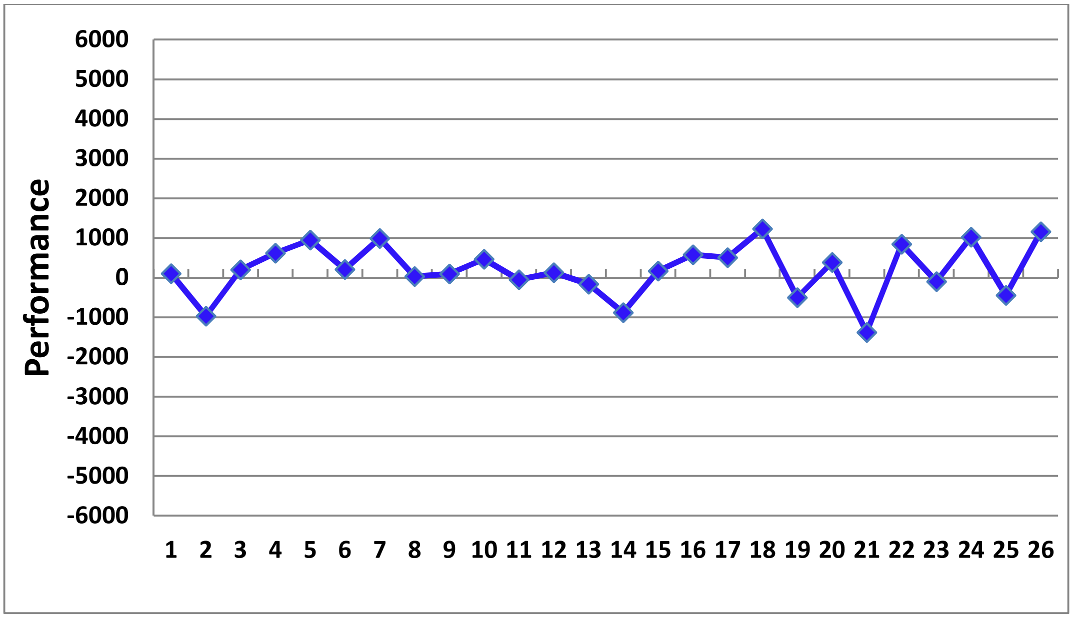

Figure A-15. Average, high tremors, no swell

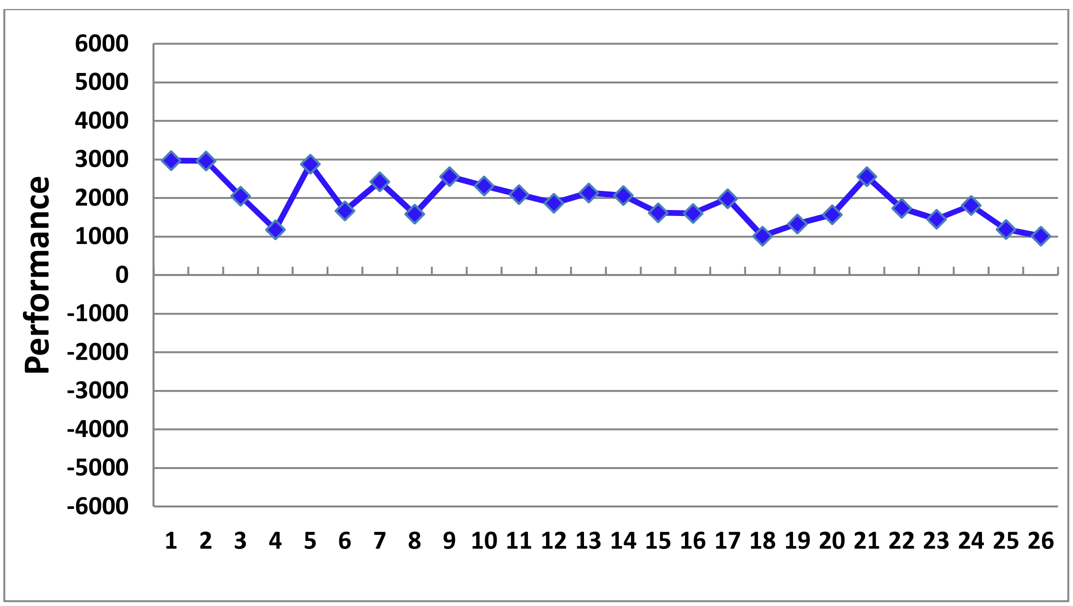

Figure A-16. Above average, high tremors, no swell

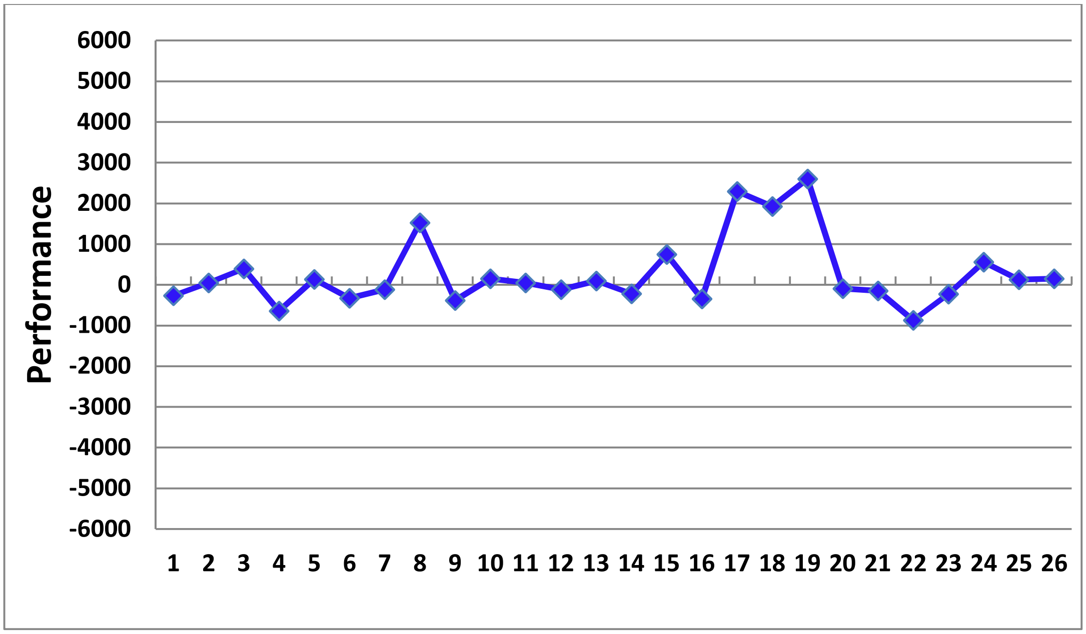

Figure A-17. Average, high tremors, positive swell

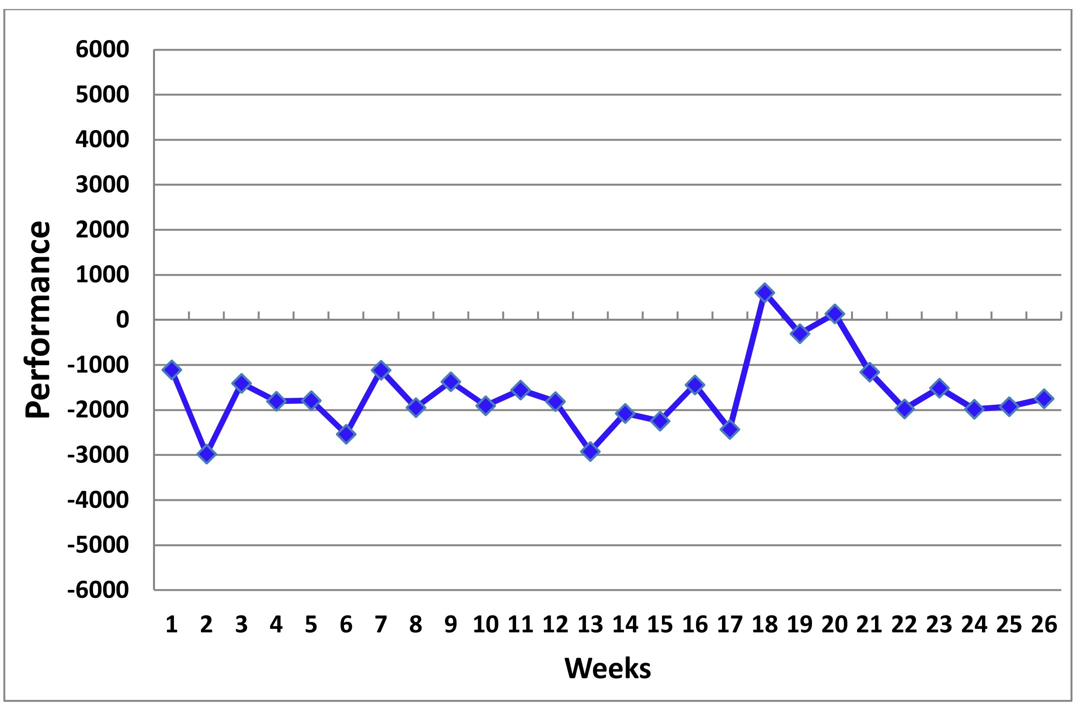

Figure A-18. Below average, high tremors, positive swell

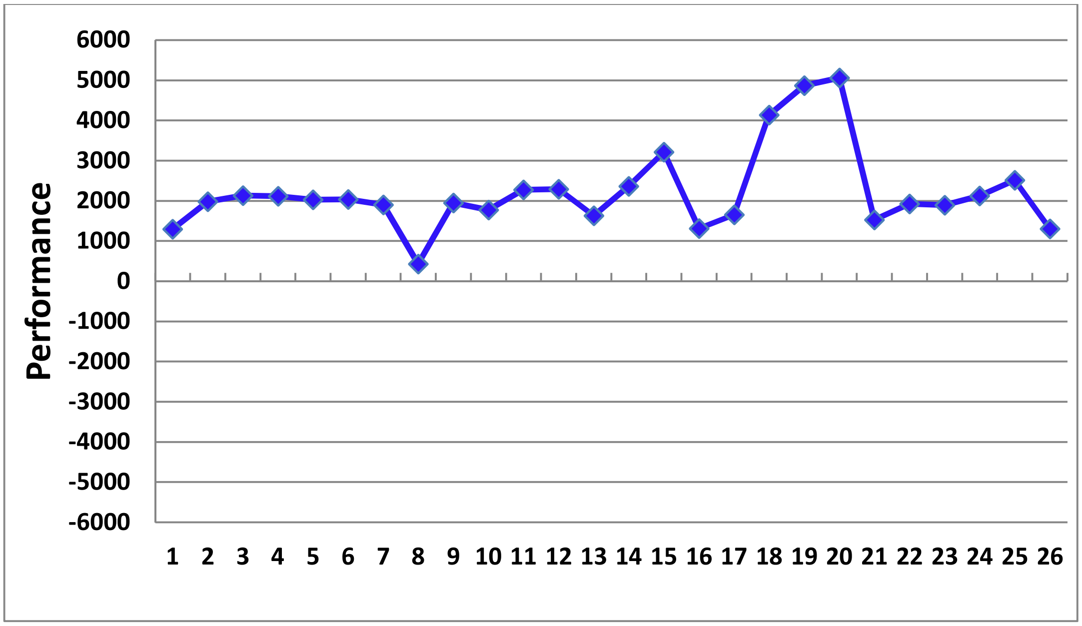

Figure A-19. Above average, high tremors, positive swell