2006

An investigation of the effects of perceived feedback accuracy on performance

Ashley A. Gray
University of South Florida

An Investigation of the Effects of Perceived Feedback Accuracy on Performance

by

Ashley A. Gray

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Edward L. Levine, Ph.D.
Walter C. Borman, Ph.D.
Douglas Nelson, Ph.D.

Date of Approval:
June 30, 2006

Keywords: performance appraisal, positive/negative distortion, perception, self-assessment, performance ratings, performance feedback

Dedication

This master's thesis is dedicated to the professors and music teachers who taught me the value of constructive criticism and performance appraisal, and to my parents, Karola Gray and Robert Grayson, whose unrelenting love and support encouraged me to reach for the stars.

Acknowledgments

I would like to thank my major professor, Dr. Ed Levine, for his guidance, availability, and consideration during the scope of this project. The following persons also deserve acknowledgement for their contribution to the work presented here:

Chavely Iglesias

Dr. Doug Nelson

Dr. Wally Borman

Dr. Mike Brannick

Table of Contents

List of Tables

List of Figures

An Investigation of the Effects of Perceived Feedback Accuracy on Performance

Ashley A. Gray

ABSTRACT

This study built upon the theoretical feedback process of Ilgen, Fisher and Taylor (1979), as refined by Kinicki, Prussia, Wu, & McKee-Ryan (2004), to contribute to a more comprehensive understanding of the mechanism underlying an individual's response to performance feedback. The feedback model implicates *source credibility* and the *feedback-rich environment* as critical elements in the process explaining recipient accuracy perceptions and responses. Thus the sign and perceived accuracy of performance feedback were investigated in a 3 X 3 (plus control) experimental design. One hundred fifty-six undergraduate students were randomly assigned to feedback conditions, in which they performed on a three-trial Lego model reproduction task. After trial one, participants received false feedback from a confederate supervisor. The feedback was based on fictional norm tables, which framed the participant's performance as falling into the 70^{th} percentile (positive), 50^{th} percentile (average), or 30^{th} percentile (negative) according to condition. The supervisor then supplemented the norm tables with conclusive comments designed to be perceived as positive distortion of the norm table feedback, reinforcement of norm tables (accurate feedback), or negative distortion of the norm tables. Performance time, errors, task-specific self-efficacy, self-assessment of performance, and self-report effort-expended data were collected on trials one through three. The results indicated a significant interaction between feedback sign and perceived feedback accuracy on participant performance ($F_{(4,132)} = 3.72$, $p < 0.01$), whereby the performance in the positively distorted-positive sign feedback condition was significantly higher than performance in the positively distorted-negative sign feedback condition. When the feedback was perceived to be accurate, negative sign feedback resulted in significantly better performance compared to the positive feedback condition (contrary to previous research). Task-specific self-efficacy was not found to mediate the relationship between type of feedback and performance, and no significant effect of feedback sign or perceived feedback accuracy on task-specific self-efficacy was found. These findings provide possible explanation as to why supervisors tend to positively distort both sign and accuracy in performance appraisals (e.g., Benedict & Levine, 1988). Implications for theoretical expansion of the feedback process model, and application to workplace performance management are discussed.

Introduction

"The thing about performance, even if it's only an illusion, is that it is a celebration of the fact that we do contain within ourselves infinite possibilities."

-- Daniel Day Lewis (February 8, 1990)

Managing Performance

The concept of performance management involves monitoring and evaluating employee performance, engaging in feedback discussions, and identifying ways to improve performance (London, 2003). These important processes are not naturally occurring phenomena, therefore it is necessary to cultivate and encourage performance management in organizations (Hillman, Schwandt & Bartz, 1990). To this end, performance appraisal and feedback are indispensable tools. Performance appraisal (PA) involves the intentional monitoring and rating of employee performance (both in terms of their current effectiveness, and their potential for improvement). It facilitates measurement of organizational performance and guides personnel decisions, but a key value of PA lies in its ability to motivate employee performance through feedback. The communication and interpretation of PA results, known as performance feedback, is an important factor in self-correcting effort and performance (Lindsley, Brass & Thomas, 1995).

Definition and Functions of Feedback.

There are four common definitions of performance feedback: 1) information received regarding the quantity or quality of one's past general performance; 2) information provided to a person following a particular performance; 3) information indicating to the performer what or how well he is doing; and 4) information about performance that enables an individual to adjust his or her performance (Alvero, Bucklin & Austin, 2001). As illustrated by these definitions, performance feedback has a multipurpose role which allows application in numerous situations to benefit both employee and organization. Feedback can provide the functional information to help an employee learn new job tasks, become familiar with work environment norms, identify opportunities to develop his or her skills and abilities, or maintain productivity (London, 2003). The communication of PA feedback can also operate in a motivational sense by channeling employee effort, helping to set reasonable performance standards/expectations, alerting the employee to his or her strengths and weaknesses, and encouraging the employee to detect errors on his or her own (Ilgen, Fisher & Taylor, 1979; Larson, 1984; London, 1988; Nadler, 1979). Practitioners and researchers have struggled to identify which characteristics of feedback and the feedback process consistently maximize its effectiveness. In the interest of addressing the issue, this paper reviews the theoretical underpinnings of a feedback process model and presents a research study in which some of the model's basic propositions are experimentally explored.

Theoretical Framework

From performance feedback literature emerge two main research themes: The first of these themes addresses the relationship between an individual's response to feedback and certain feedback characteristics such as *source credibility*, feedback sign, specificity, and consistency, as well as the acceptance of feedback. A second theme is centered on the potential antecedents of feedback-seeking behaviors. Research conducted on the first set of topics has been criticized for ignoring important mechanisms which trigger and perpetuate the feedback process (Kinicki, Prussia, & McKee-Ryan 2004; Fedor, Davis, Maslyn, & Mathieson, 2001). This argument is duly noted in much of the empirical work to date which has focused primarily on the surface relationships between these feedback characteristics. Cognitive processes and perceptions, though less visible than behavior, represent important components of any model of human behavior. Therefore, it is the intention of the present study to contribute to a more comprehensive understanding of the mechanism underlying an individual's response to feedback.

The feedback model proposed by Ilgen, et al (1979), and refined by Kinicki et al (2004), provides a worthy theoretical vehicle to guide both the integration of empirical findings and the development of meaningful research hypotheses. Illustrated in Figure one, the model is predicated on the notion that certain psychological or cognitive processes may activate different or inconsistent responses to feedback, which are often observed in feedback research (Kluger & DeNisi, 1996). *Feedback-rich environments* (large amounts of specific and positive feedback) and *source credibility* (as perceived by the employee) are proposed as antecedents to performance. The influences of the antecedents are mediated by three cognitively-oriented constructs: *perceived accuracy*, *desire to respond* to feedback, and *intended response*. *Perceived accuracy* represents whether or not the feedback recipient views the feedback as accurately representing his or her performance, which can be likened to "acceptance" of feedback. *Desire to respond*

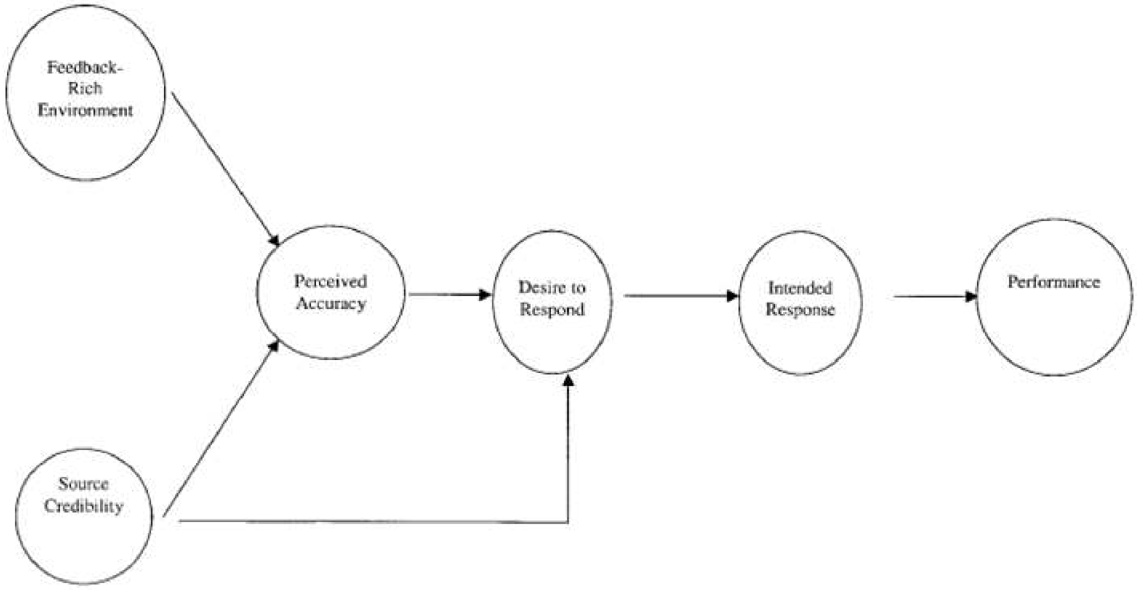


Figure 1. Model of feedback process (Ilgen, et al., 1979; Kinicki, et al., 2004).

and *intended response* represent motivational aspects of the process, influenced by *perceived accuracy* of feedback and affecting recipient *performance*.

Framing Perceived Accuracy

It is necessary to diverge from discussion of the model to expand upon the concept of *perceived accuracy* as it relates to self-assessments. Formal self-assessment (SA) is a procedure commonly studied with feedback and used in organizations (Steele & Ovalle, 1984) whereby a person evaluates his own performance. The literature identifies numerous benefits of SA such as helping employees understand their work environment and performance expectations (London, 2003), encouraging self-regulation, and reducing the ambiguity of performance expectations (Ashford, 1989). The role of SA in influencing the *perceived accuracy* of feedback is based on the accuracy of the SA itself, which is typically defined as the congruency between self and supervisor ratings (Alvarez & Bernal, 1999). There is some substantiation that people are capable of making objectively accurate ratings of their own performance, as evidenced in Greller & Parson's (1992) study in which task feedback was more strongly associated with self-rating than was feedback from other persons. However, in the absence of clear and easily-accessible objective task feedback, SAs suffer from positive bias, especially regarding quantity of work and ability for the job (Alvarez & Bernal, 1999).

SA is not always directly observable, thus it is frequently overlooked as a factor in the feedback process (unless a formal SA procedure is involved). It is always important to consider the effects of SA as people are likely engaging in a form of self-appraisal in daily life. Considering that a person is likely to perceive his or her own SA as accurate and that people tend to overrate themselves, there are some obvious implications for how SA might interact with the *perceived accuracy* and sign of feedback from other sources. For example, greater incongruencies between feedback provided by an external source and SA may be likely to be result in perceptions of inaccurate feedback, which lowers the *desire to respond* to the feedback. Similarly, a formal SA might strengthen a self-opinion that causes an employee to be more resistant to feedback from others (Blakely, 1993). On the other hand, discrepancies between self-evaluation and evaluation by others may also signal the need for performance improvement and subsequent action toward that end. A discrepancy which occurs through the receipt of negative feedback has been shown to, at a minimum, decrease the likelihood of overrating oneself (Atwater, Rousch & Fischthal, 1995).

Feedback-Rich Environments

A number of factors contribute to the existence of a *feedback-rich environment*, including: feedback focus, specificity, sign, and source. As implicated by the feedback process model, feedback characteristics are proposed to influence performance through *perceived accuracy.* This mechanism offers a useful theoretical explanation for the relationships between these variables and recipient performance, even though it is rarely (if ever) acknowledged.

Behavioral focus and specificity. Feedback targeting behavior consistently yields higher performance (Meyer, Kay & French, 1965) and motivation for improvement (Burke, Weitzel, & Weir, 1978) as compared to personality-focused feedback. In instances involving negative feedback it is more difficult for a recipient to reject or deny objectively-measured performance feedback versus a subjective opinion (London, 2003). Behavior is more amenable to quantification than is personality, thus its evaluation is more likely to be accepted or perceived as accurate. In a similar vein, feedback specificity has also proved beneficial in the literature (Russell & Goode, 1988). The more specific the communicated information and the extent to which it reflects the nature of the task, the more compelling, easy to understand, and less threatening it is when compared to descriptions of personal characteristics (Kluger & DeNisi, 1996). Specific feedback is likely to be perceived as more factual and perhaps objective, than general or vague feedback which is likely to be seen as unfounded.

Sign. Feedback sign behaves as the "juicy" and controversial characteristic in the feedback family, both concerning recipient perception and response and in research findings. One difficult issue is the way in which feedback sign is defined; feedback sign may be absolutely defined (e.g., positive or negative assessment compared to a standard or objective, Kluger & DeNisi, 1996) or relatively defined (e.g., positive or negative in relation to a recipient's self-appraisal). Research is not consistent in the specification of sign definition, as seen in highly interactive findings that are observed across the literature. Despite this inconsistency, the strong influence of sign is generally acknowledged. Negative feedback has been shown to hurt performance when the recipient's personal characteristics are targeted (DeNisi & Kluger, 2000), and when it is behaviorally-based (Ilgen, et al, 1979; Ashford, 1989). Interestingly, when negative feedback is provided the *perceived accuracy* of feedback tends to be quite low, and it may generate anger and discouragement (Brett & Atwater, 2001). By contrast, positive feedback tends to yield better outcomes, relative to negative feedback, in terms of motivating performance (Parsons, Reid, & Crow, 2003), producing better relationships, more pleasant emotions and constructive behavioral intentions (Van de Vliert, Shi, Sanders, Wang, & Huang, 2004). Theoretically, positive feedback is more likely to be perceived as accurate, regardless of behavioral vs. personal focus or specificity because it does not indicate a deficiency that requires action. Self-enhancement theory (Shrauger, 1975) suggests that people will be more open to receiving positive feedback than negative feedback because they want to confirm a positive image of themselves, while negative feedback is often found to be threatening and subsequently rejected (Kluger & DeNisi, 1996). These considerations may cultivate a lay belief in the need to "accentuate the positive" among supervisors, and could explain the strong tendency of raters to positively distort both positive and negative feedback as found by Benedict & Levine (1988). As will be discussed later, this acceptance of the feedback should result in the *desire to respond*, thereby yielding positive performance outcomes.

Source Credibility

In addition to specific and positive feedback aspects outlined by the model's *feedback-rich environment*, feedback *source credibility* is also posited to affect perceived feedback accuracy. As noted in the discussion of SA, feedback can be obtained from the self, other people, or directly from a task. Presumably, feedback from the self is typically deemed "credible." When feedback is provided by another source, the issue of credibility becomes more variable. *Source credibility* has been shown to affect recipient intention to use feedback while favorably influencing recipient evaluations of the source and the feedback (Bannister, 1986; Suzuki, 1978; Albright & Levy, 1995). The Ilgen et al. and Kinicki et al. model suggests that *perceived accuracy* will determine the relationship between *source credibility* and intention to use the feedback or *desire to respond* and *intended response*. Several researchers have reported an interaction of effects between *source credibility* and feedback sign on performance (e.g., Podsakoff & Farh, 1989; Fedor, et al., 2001; Steelman & Rutkowski, 2004). Interestingly, it appears that *source credibility* can not only cancel the unfavorable effects of negative feedback but it can actually trigger performance improvement in scenarios involving negative feedback. At a basic level, negative feedback from a credible source results in better performance than negative feedback from a less credible source. The same relationship, however, does not necessarily exist in conditions wherein credibility is manipulated with positive feedback (Podsakoff & Farh, 1989). This interaction may be related to the theoretical model whereby positive feedback is part of the *feedback-rich environment* and is therefore tightly linked to *perceived accuracy*. Negative feedback is not an element of the theoretical model's *feedback-rich environment* and is therefore its *perceived accuracy* is dependant, in part, on *source credibility*.

To summarize the model's antecedents and *perceived accuracy* variable, the characteristics of a *feedback-rich environment* and *source credibility* represent theoretically supported relationships in the literature. In addition, Kinicki et al.'s (2004) test of the complete model paths demonstrate a positive relationship between the two antecedents and *perceived accuracy*. *Source credibility* also exhibited a positive relationship with the *desire to respond* to feedback, indicating that it wields some level of direct influence over a recipient's interest in responding (independent of *perceived accuracy*).

Cognitive Components

Desire to respond and intended response. The model next highlights the importance of acceptance of feedback in predicting a recipient's *desire to respond* and his or her *intended response*. In particular, perceived feedback accuracy is posited to override feedback sign such that high *perceived accuracy* should result in a positive *desire to respond* when compared to perceived inaccuracy, irrespective of feedback sign. Stemming from this *desire to respond* are behavioral intentions which are magnified when the feedback is associated with a future outcome. Kinicki et al (2004) provided the first investigation of the relationship between *perceived accuracy* and the *desire to respond*, and between the *desire to respond* and *intended response*. A positive relationship was established in both cases, and between the proximal predictor of *intended response* and performance outcome.

As emphasized by Kinicki et al (2004), the mediation of cognitions represents an important attempt to dig beneath the surface of the feedback-response process. This focus on understanding perceptions and accuracy in the appraisal of feedback draws a parallel to an extensive body of PA literature which revolves around reducing rater bias and the cognitive mechanisms involved in the rating process. Raters are faced with a complex task in which they must observe, store, and recall information, integrate and judge effectiveness of behaviors, and translate judgment onto a rating scale (Steiner, Rain & Smalley, 1993). This procedure can be further complicated by time constraints, insufficient information (Barnes-Farrell, 2001), rater bias, and lack of enthusiasm (London, 2003). In the interest of obtaining a performance evaluation that reflects the employee's actual contribution to the organization, rating "accuracy" has become a top priority in performance appraisal research. To attain this goal raters are often trained to avoid any type of rating inflation or distortion (e.g., halo bias).

Self-efficacy. The overwhelming support obtained by Kinicki et al. (2004) for mediation of the feedback-performance relationship by a set of cognitive variables warrants consideration of additional cognitive variables which may operate in the underlying feedback mechanism. One potentially mediating variable between *perceived accuracy* and performance is self-efficacy. Generally defined, self-efficacy is the belief that one can "organize and execute" the behavior necessary to deal with certain situations, or to produce the required outcome (Bandura, 1980; Bandura, 1977). This construct captures a motivational self-belief about task capabilities (Chen, Gully, & Eden, 2001), and can be specific to a task rather than generalized. Self-efficacy theory (Bandura, 1977) implicates four sources of background information as antecedents to a person's development (performance accomplishments, vicarious learning, levels of emotional arousal, and social persuasion) and it indicates that self-efficacy has at least three behavioral consequences (approach vs. avoidance behavior, quality of performance, and persistence in the face of obstacles or disconfirming experiences). The "performance accomplishment" antecedent is directly relevant to the present discussion of performance feedback, while the "approach vs. avoidance and persistence despite disconfirming experiences" consequences speak to the *desire to respond* and *intent to respond* constructs in the feedback model. Research supports the buffering effects of high self-efficacy against the negative effects of negative sign feedback (e.g., Bandura, 1989; Phillips & Gully, 1997). In the case of negative sign feedback, perhaps the high self-efficacy could be a function of past performance accomplishments, and it yields a persistent "approach" or *desire to respond*-reaction to negative feedback. However, the literature also suggests that self-efficacy itself is vulnerable to negative feedback, perhaps affected through negative emotions elicited by the feedback. As a second illustration of the susceptibility of self-efficacy to feedback, Tang & Sarsfield-Baldwin (1991a) found that participants receiving false positive feedback attributed higher ability and effort to themselves and exhibited higher intrinsic motivation than those who received false-negative feedback (1991b). In the context of *perceived accuracy* it is possible that feedback perceived to be accurate may exert more of an impact on self-efficacy and the subsequent *desire to respond*. Feedback perceived as inaccurate may simply be dismissed, thereby protecting self-efficacy (especially when the sign is negative). The wide variety of theoretically-based hypotheses which may be posited from self-efficacy theory and the feedback model indicates a need for further research and theoretical development.

Limitations of Previous Research

The investigation by Kinicki et al (2004) marked the first attempt to test the widely-held assumption that the accuracy of performance appraisals and subsequent performance are positively related. One limitation of Kinicki et al's (2004) investigation of the feedback model was the use of same-source survey data. In addition, it is assumed that misleading, incomplete or inaccurate feedback would have dysfunctional consequences (London, 2003). To date no research has attempted to experimentally isolate the perception of feedback accuracy (or distortion) in order to determine the role of recipient feedback acceptance in the feedback process and consequential performance outcomes.

A small number of studies have manipulated the verity of feedback, by providing accurate and erroneous feedback for time estimation tasks (e.g., Ryan & Robey, 2002; Brosvic & Finizio, 1995) and mathematical tasks (e.g., Cummings, Schwab & Rosen, 1971). These studies typically report that accurate feedback is helpful in centering a participant's time-estimation around the target interval, or that negatively distorted feedback results in lower performance on a math task. Unfortunately, this line of research has tested only negative distortion conditions, ignoring the potential effects of positive distortion of feedback. Issues of task relevance and experimental design (e.g., unsubstantiated conclusions by Cummings et al, 1971) limit the generalization of the findings to other contexts, and these studies do not address the issue of the recipient's perception of feedback accuracy. A review of main points is presented in Table two to facilitate the integration of research findings which have addressed different components of the feedback process.

Table 1
Summary of Main Points

- ✓ Feedback is an integral part of performance management (e.g., London, 2003; Lindsley, Brass & Thomas, 1995).
- ✓ Researchers and practitioners have stressed the need for objectively accurate PA ratings and performance feedback delivery (e.g., Murphy & Cleveland, 1995). Supervisory PA ratings tend to be inflated (e.g., Benedict & Levine, 1988) and are thus the target of extensive training and research aimed at removing any positive distortion performance ratings and feedback (e.g., Maroney & Buckley, 1992; Schleicher & Day, 1998).
- ✓ Kinicki et al (2004) suggest that a feedback recipient's future performance is affected not by the absolute accuracy of feedback, but by the *perception* of feedback accuracy as these perceptions influence the recipient's *desire to respond* and his *intended response*.
- ✓ Factors contributing to higher *perceived accuracy* or *acceptance* include *feedback-rich environment* (feedback has a behavioral focus, is specific and positive) and *source credibility* (source has expertise and/or experience; Kinicki, et al., 2004)
- ✓ The recipient is also likely to determine the accuracy of feedback (from an external source) through comparisons with self-opinions or self-assessments of his performance (Alvarez & Bernal, 1999).
- ✓ The more similar the self- and other-evaluations are, the higher the perceived accuracy of the feedback from an external source. However, self-ratings tend to suffer from positive bias (Alvarez & Bernal, 1999), increasing the likelihood that external feedback will be perceived as inaccurate.
- ✓ No experimental evidence exists to support the assumption that performance following delivery of <u>objectively</u> accurate feedback is any better than that which follows the delivery of inflated feedback.
- ✓ Research has not experimentally manipulated perceived feedback accuracy to determine the effects of <u>perceived</u> feedback distortion or accuracy on subsequent performance.

The Present Study

The empirical work to date has primarily focused on surface relationships between feedback characteristics. Cognitive processes and perceptions, while less visible than behavior, represent important components of any model of human behavior. The negative effects of feedback present a challenge for organizations as the dissemination of negative feedback is often necessary to address poor employee performance in the workplace. Therefore one important issue is to discover a method which facilitates the effectiveness of feedback for poor performance (Goodman & Wood, 2004). It was the intention of the present research to contribute to a fuller understanding of the mechanism underlying an individual's response to feedback, where the response is measured by subsequent performance. This objective was addressed by an investigation of the capacity of *perceived accuracy* of feedback to result in higher performance compared to feedback perceived to be distorted. It is important to reiterate that the mechanism of interest in the present study does not involve the absolute accuracy of performance feedback, but rather the recipient's perception of feedback accuracy.

In addition, several benefits of positive feedback have been discovered and replicated, and evidence supports the performance-impeding effects of negative feedback. The literature suggests that feedback sign is likely to interact with *perceived accuracy* (e.g., Brett & Atwater, 2001), thus perceived feedback sign and *perceived accuracy* were experimentally manipulated in the present study in order to examine the specific nature of the potential feedback sign**perceived accuracy* interaction. Such information would be instrumental in furthering the development of feedback process theory. Ultimately, if the perception feedback being accurate does not lead to better performance in any context (i.e. when feedback is positive or negative), this knowledge would have far-reaching implications for future performance appraisal accuracy research, rater training research, as well as application in the workplace.

The design of the present study was guided and inspired by the feedback process model (Ilgen, et al., 1979; Kinicki, et al., 2004), whereby the recipient's *perceived accuracy* was manipulated by varying the *feedback-rich environment* (feedback sign) and *source credibility* (supervisor credibility). In an attempt to better understand the cognitive processes involved in the feedback-response process, self-assessments and self-report ratings of effort expended were gathered. The self-assessments could potentially provide a deeper understanding of the recipient's feedback accuracy perceptions. Similarly, recipient effort-expended should provide an "inside look" at *intended responses* to the feedback to complement the performance measure as outcome of interest.

Study Hypotheses

H1. Provision of feedback yields higher performance in comparison to withholding feedback.

H2. There is a main effect of perceived feedback accuracy, such that accurate feedback will result in significantly better subsequent task performance than distorted feedback. Feedback perceived as accurate will encourage higher performance relative to feedback perceived as either positively or negatively distorted (Kinicki, et al, 2004). This hypothesis addresses the cognitive aspect of a widely held assumption upon which most performance appraisal and rater training research is based: If an employee receives inaccurate or distorted performance feedback, his development is presumably hindered and performance is unlikely to improve.

H3. There is a main effect of feedback sign.

H3a. Feedback with a positive sign will result in significantly higher performance than that resulting from negative sign feedback. Positive feedback should motivate performance (Parsons, et al., 2003) and produce constructive behavioral intentions (Van de Vliert, et al., 2004). While negative feedback reveals the need for performance improvement, it is likely to be perceived as inaccurate and discouraging (Brett & Atwater, 2001), thus damaging performance.

H3b. There are no differences in performance effects between positive sign and average sign feedback. Recipients of average feedback will recognize the need to improve their performance and they are less likely to dismiss the feedback or be discouraged, as can occur with negative feedback. Positive feedback may result in slightly higher performance, but it is not expected to be significantly different from the performance associated with average feedback.

H4. Feedback accuracy and sign interact to affect performance.

H4a. When feedback is positively distorted, positive sign feedback will result in lower performance than positively distorted-negative feedback. Participants who perceive positive feedback to be positively distorted instead of accurate are not likely to recognize any need to improve performance (affecting *their desire to respond*), and thus will not exhibit higher levels of subsequent performance. Conversely, the inaccuracy perceived in positively distorted feedback may be encouraging or motivating to participants receiving negative feedback, resulting in higher performance than does positive distortion of positive feedback.

H4b. When feedback is negatively distorted, negative sign feedback results in significantly lower performance than average sign or positive sign feedback. The negative distortion of feedback which already has a negative sign may be so discouraging to recipients that they feel they can't improve, and may lead to unimproved subsequent performance.

H4c. When feedback is perceived as accurate, positive sign feedback yields significantly higher performance than either average sign feedback or negative sign feedback. Feedback perceived as accurate also results in significantly higher performance when the sign is average, as compared to when the sign is negative. This proposition is similar to that of the main effect for feedback accuracy because, unless otherwise instructed, the participant is expected to assume the feedback is accurate and accurate feedback is expected to enhance performance except when performance is already positive.

*Exploratory***.** There is also interest in: 1) The role of self-efficacy as a mediator between *perceived accuracy* of feedback and performance, and 2) the effects of *perceived accuracy* of feedback on recipient self-efficacy. Due to the lack of theoretical suppositions about how self-efficacy might operate in a recipient's response to feedback, self-efficacy was investigated on a purely exploratory basis.

Method

Sample

Participants in this study were 206 undergraduates from a large southeastern university. Forty-two of these participants participated in the experiment pilot and are not included in the final study sample. Additionally, the data for eight participants was discarded due to error in procedures during the experimental sessions (e.g., supervisor failed to collect the necessary performance data, failed to correctly apply experimental condition, or did not time the session properly). The final study sample was comprised of 156 participants, ranging in age from 18 to 50 years (M = 21.16 years). Seventy-four percent of the sample was female, and a variety of ethnicities were represented (57.7% White, 17.9% Hispanic/Latino, 14.1% African-American, 5.8% Asian, 4.4% Other or not reporting). Of the 97 participants who reported being employed, 79% worked part-time. When asked how often they had played with Legos or other building sets, 9.6% indicated "very often," 10.3% responded "often," 17.3% chose "sometimes," and 62.8% answered "rarely" or "never."

Designing Performance Feedback

Research has indicated that feedback delivery by managers or supervisors yields the most consistent performance effects, when compared to other human sources (Alvero et al, 2001). Graph-form data accompanied by either written or verbal feedback yields the most consistent effects in improvement, and the use of a standard performance referent as a basis for comparison in appraisal feedback is also recommended (Alvero, et al., 2001). These findings were taken into consideration in the design of feedback for the present experiment. Norm tables which depicted percentile rank were selected as the source of feedback, were crafted to appear as though they were "Lego-issued," and provided a visual scale of percentiles, to facilitate comprehension of where certain percentiles fall in the distribution. The norm tables were accompanied by verbal comments from the confederate experiment "supervisor." Five undergraduate students were recruited to act as confederate supervisors. Each supervisor was blind to the study hypotheses and completed a structured two-week training course which provided specific instruction and practice for running the experimental sessions and delivering the condition-appropriate feedback.

Manipulations

Design. The study employed a 3 X 3 (*perceived accuracy* x sign) between-subjects factorial design as specified in Table two. A tenth condition served as a control group in which participants received no feedback. Participants were randomly assigned to the ten conditions.

Table 2
Experimental Conditions

Sign	Perceive Feedback as Accurate	Perceive Feedback as Positively Distorted	Perceive Feedback as Negatively Distorted
Positive (70%)	Condition 1	Condition 2	Condition 3
Average (50%)	Condition 4	Condition 5	Condition 6
Negative (30%)	Condition 7	Condition 8	Condition 9

Sign. Feedback sign was manipulated through the use of false percentiles in order to convince the participant that his/her performance had been of a certain quality. Three sets of tables were created to portray the participant's completion time as falling into the 70^{th} percentile (Positive sign condition), the 50^{th} percentile (Average sign), or the 30^{th} percentile (Negative sign). These levels were selected so as to create clear differences among conditions and to avoid extreme positive or negative levels. In each condition, tables were created for all possible task completion times in five second increments ranging from one minute to 20 minutes (see Appendices A, B and C). During feedback administration, the participant's performance was "evaluated" according to a condition-appropriate table which matched the participant's completion time. The norm table was introduced in the following manner, "This is a LEGO norm table for the model you just completed. It gives you your percentile rank—basically it tells you what percentage of people had a completion time longer than yours—or in other words, what percentage of people you performed better than. The table uses numbers a little differently from grading for a college course, so for example, in a class somewhere around 70 might be the median score, meaning that ½ the class was higher than 70, and ½ the class was lower than 70. Here, the 50 is the percentile level for the average performance time. Your time was ______" *(show his/her time on the timer).* We tend to round to the nearest 5 seconds, so the table clearly indicates that your performance time falls into the __th percentile of performance, which is [above/below] average. So this means that__% of people did better than you and __% of people did worse than you. The table provided by LEGO for this model is based on the performance of hundreds of students and is very accurate."

Accuracy. Accuracy was not operationalized as reflecting some true score for performance. Instead the accuracy manipulations were intended to create a perception or a presumptive "accuracy" for the recipient (which may or may not be accurate as a true score). In other words, the purpose of the accuracy manipulation was to produce a feedback discrepancy which would be perceived by the participant as inaccuracy or distortion on the part of the supervisor. The confederate supervisor administered bogus feedback after a participant had completed the first of three performance trials. (Participants in the control condition did not receive performance feedback of any kind.) The initial feedback medium was a bogus norm table which indicated the participant's "percentile rank" for performance on the task. The participant's perception of the feedback as being accurate or positively/negatively distorted was then manipulated by variations in the supervisor's comments accompanying the norm table presentation. Specifically, the distortion conditions required the supervisor to verbally conclude that the norm table was either too low (positively distorted feedback) or too high by 15 percent (negatively distorted feedback), while in the accurate condition the supervisor agreed with the norm table and provided specific reasons for her "conclusion." The comments assigned to each condition are presented in Appendix D. The following is an example of a comment which was used to apply the positive distortion manipulation, "(*After explanation of the participant's performance according to the Lego norm table)* I bet you probably should be rated higher, I kinda liked what you did, I'm willing to say you're probably closer to the __% *(add 15%)* percentile…but then again, the other supervisors in this lab say I'm too lenient and that I give people too high an impression of their performance. Still that is what *I* think."

Bonding Opportunity

One criticism of laboratory experiments designed to explore the impact of the rater/ratee relationship and PA processes is that the participants don't actually have any contact with the confederate prior to or following the study. Studies have shown that supervisor relationships have a significant impact on the feedback and social support perceived by employees (Kramer, 1995). Benedict & Levine (1988) improved upon previous research by ensuring that the participant raters had previous personal contact with the "subordinate" to be rated, so that supervisors could interact with the subordinate and form some sort of bond with the subordinate, as well as learn the consequences of the ratings to be given. The present experiment employed a method similar to that of Benedict & Levine (1988) whereby subjects had the opportunity to bond with the confederate supervisor before "beginning" the experiment. The purpose of this bonding component was twofold: 1) To increase relevance to the real-world, in which employees rarely receive feedback from a complete stranger, and 2) To encourage a sense of responsibility or worth in the participant regarding his/her performance in the experiment. The supervisor bonding script (Appendix E) required the supervisor to solicit help from the participant to complete a short paperwork task "unrelated to the experiment." The script also suggested several specific questions intended to elicit conversation from the participant.

Incentive

Incorporated into the experimental design were several attempts to improve the relevance of good task performance to the participants. In order to encourage the effort expended on the task, participants were told that performance on the task is an indicator of fundamental and important cognitive and psychomotor abilities. Compensation was offered in the form of extra points (1 point per half hour) credited toward the student's psychology class. Participants were also informed that a reward of $20.00 gift certificates would be offered to the four highest performers in the experiment. Finally, a series of Lego models were chosen to serve as the performance task for the experiment. It was expected that such a task might be fun or interesting to undergraduate participants, thus fostering additional attention to their performance.

Task

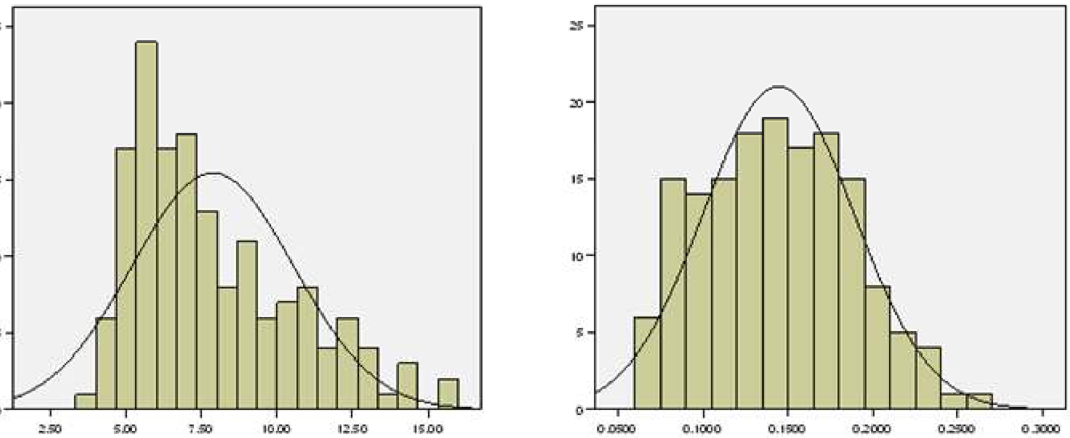
In addition to its potential to foster participant interest, a Lego model-reproduction was selected as an appropriate task for three reasons: 1) Successful performance on the task is achieved by integrating cognitive ability and psychomotor skills (more complex relative to the time-estimation and math-problem tasks used in previous research); 2) The difficulty is easily customized by increasing the number of extra pieces from which the participant must choose, and by using models which have adjacent pieces of the same color; 3) Participant performance can be easily (and reliably) quantified by recording completion time as well as number and type of errors committed, providing a well-rounded representation of performance.

As applied in the experiment, the task consisted of three Lego model reproduction trials. During each trial, the participant was instructed to reproduce a model by selecting the correct pieces from a pile and assembling them without touching or taking apart the sample model. The goal of the task was to complete the model in the shortest amount of time possible, and the model reproductions were required to match both color and piece size. The performance measures collected included model completion time and number of errors in the reproduction. In order to control experiment length, each trial length was limited to 13 minutes. Model building was halted if 13 minutes had elapsed and the participant was still working. The exact task instructions are available in Appendix F.

Measures

Performance. Completion time and number of errors were recorded for each model that was built. Errors included incorrect colors, piece sizes, or pieces missing when 13 minutes had elapsed. For each model, the number of errors was multiplied by five seconds and then appended to the model completion time. The primary performance variable used for hypothesis testing was computed by taking the average completion time (in minutes) for models two and three (the two models following administration of performance feedback). The correlation for performance across the two models was 0.71. The Spearman-Brown formula was used to correct for doubled length, and indicated a reasonably good reliability for the performance means ($r = 0.83$). In eight experimental sessions the supervisor forgot to stop the participant at 13 minutes. To facilitate application of the same time and error rule to all participants, the average number of seconds per piece assembled was calculated. Based on the average of 20 seconds per piece, these 8 participants were penalized one error (on top of any other errors they had committed) for every 20 seconds they had been allowed to continue beyond 13 minutes. This formula was intended to simulate the number of missing pieces which would have likely existed at 13 minutes.

Examination of the performance distribution revealed that the data were positively skewed (Fisher's $\gamma = 0.83$, standard error of skewness = 0.19). Therefore, a reciprocal transformation was applied, which noticeably improved the skew of the distribution (Fisher's $\gamma = 0.25$, standard error of skewness = 0.19). A side-by-side comparison of the untransformed and transformed performance distributions is presented in Figure 2.



Untransformed performance distribution

Transformed performance distribution

Figure 2. Side-by-side comparison of performance outcome distributions.

Self-assessment and effort rating. The picture of participant performance painted by the objective performance measure just described was augmented by the inclusion of two single-item self-report scales which were administered following the completion of each model. These items were designed to measure how the participant viewed his/her own performance. First, the amount of effort expended on the previous trial was measured as participants responded to the Effort Rating Scale (Appendix G). Second, participants evaluated their own performance using the Simple Self-Assessment scale, in which a participant would indicate the level of their performance in percentile form (Appendix H).

Self-efficacy. Task-specific self-efficacy (TSSE) was assessed with a format similar to that employed by Cannon-Bowers (1988) and Locke, Frederick, Lee, and Bobko (1984). The scale design was based on Bandura's recommendations of rank ordering a number of tasks according to difficulty, to which the participant would indicate (yes/no) whether he believed he could complete each one. For each positive response provided the respondent indicated his degree of certainty (0% to 100%) that he could complete that particular task. The magnitude of the participant's self-efficacy was measured by the number of "yes" responses, and the strength of his/her self-efficacy was assessed by dividing the certainty response (%) total by the number of "yes" responses. A total combination self-efficacy score was calculated by summing the magnitude and strength totals. This combination was advantageous because it weighted magnitude and strength equally. The TSSE scale (Appendix I) consisted of two sets of task statements, one targeted completion time and the other addressed errors. With the exception of the practice model, participants completed both sections of the TSSE while they viewed the upcoming model. The TSSE for each trial was obtained by taking the average of the completion time TSSE and error TSSE. These four measures of TSSE comprised the overall TSSE, which demonstrated reasonably good reliability (Spearman-Brown r = 0.87).

Participant Standing Sheet (PSS) and Ability Study Questionnaire (ASQ). The Participant Standing Sheet (Appendix J) was included to serve as a manipulation check for feedback accuracy. Participants completed the PSS under the impression that it was included as a "quality control" component assessing the accuracy of the supervisor's feedback. The Ability Study Questionnaire (ASQ) displayed in Appendix K was administered at the conclusion of the experiment. Participant perceptions were measured by questions addressing the participant's performance in trial one, and repeating the evaluation of the accuracy of the supervisor's feedback. Additional questions gauged the participant's previous experience with Legos or similar building tasks, and requested demographic information.

Pilot Testing

Forty-two pilot sessions were conducted to develop and refine the task, experimental manipulations, and measures. "Pilot sessions" also included experiments in which supervisors received training. Initial testing of the task revealed that undergraduates perceived it to be relatively interesting. Two models were replaced when the pilot data suggested that one model was too simple, and another was too difficult to finish in a timely manner. The reward incentive was also modified to include a choice of gift certificates to improve participant interest. Piloting provided guidance as to what anchors would be appropriate for the TSSE scale and norm table "filler" percentiles. Pilot participants were questioned to determine whether the hypotheses had been suspected during the experiment. Two participants reported guessing that the experiment was about feedback, but no participant responses suggested that the hypotheses had been compromised in the experiment. Finally, it was determined that the accuracy manipulations needed to be more salient, therefore the level of distortion was raised by five percent in either direction (i.e., instead of 10%, the distortion comments would add 15% to, or subtract 15% from, the levels specified by the sign conditions described earlier), and the distortion comments were rewritten to be more "flimsy" (as written in Appendix D).

Procedure

Students were recruited to participate in the experiment for one hour, with only one participant assigned to each time slot. The experimental session began when the arriving participant was greeted by a confederate (author) and asked to take a seat while the "supervisor" (research assistant) finished some material she was "trying to finish."

The confederate introduced the supervisor as having had considerable training on the task that the participant would be performing. The confederate verbally confirmed that the supervisor had "all the copies she needed" and excused herself. Moments later the supervisor began the scripted bonding opportunity (Appendix E) while continuing to appear busy, and she attempted to engage the participant in a conversation which involved casual (and non-threatening) questions. During this period the supervisor also solicited help from the participant in finishing her task "before beginning the experiment." The conversation and helping-task time was recorded. Next, informed consent was administered, followed by the Lego task instructions (Appendix F) and presentation of a practice model to build.

After completion of the practice model, the first official model was placed on the table before beginning the first trial, and the participant completed the first TSSE scale. The participant was then allowed to work on the first model. Upon completion of trial 1, the supervisor administered the first Effort Rating and Self-Assessment scales, while looking up the condition-appropriate norm table. The supervisor collected the two scales and provided the scripted performance feedback (Appendix D). Feedback was not provided in any form during the remainder of the experiment.

Following the performance feedback, two additional models were replicated, each preceded by administration of the TSSE scale and trailed by and completion of the Effort Rating and Self-Assessment. After completing trial three the participant was asked to fill out one more TSSE (under the impression that he/she would have to build model 4) as well as the ASQ. After the ASQ was completed, the participant was told that it was not necessary to build model 4 because the experiment was "actually over." The supervisor debriefed the participants and thanked them for their participation.

Results

Manipulation Check

Participant responses to two items were analyzed to determine the success of the feedback sign and feedback distortion experimental manipulations. First, responses to Ability Study Questionnaire item one, "*In my opinion, my performance in trial 1 (Robot) was: 1) Below Average (Below 50th percentile); 2) Average (50th percentile); 3) Above Average (Above 50th percentile)*" were subjected to a 3 X 3 analysis of variance to determine the effect of the feedback sign manipulations. The results reported in Table three support a main effect for feedback sign ($F_{(2, 132)} = 47.77, p < .01$), but no main effect of feedback accuracy ($F_{(2, 132)} = 0.27, p = 0.363$) or sign*accuracy interaction ($F_{(4, 132)} = 0.490, p = 0.128$). Tukey HSD post hoc tests revealed a significant difference ($p < 0.01$) between all levels of feedback sign, suggesting a successful manipulation of

Table 3
ANOVA Results for Feedback Sign Manipulation Check

	Sum of Squares	*df*	Mean Square	*F*	*p*
Intercept	692.20	1	692.20	2580.34	0.00
Sign	25.62	2	12.81	47.76	0.00
Accuracy	0.45	2	0.23	0.84	0.435
Sign*Accuracy	1.79	4	0.45	1.67	0.16
Error	35.41	132	0.27		
Total	749.00	141			

Note. Dependent variable is participant response to Ability Study Questionnaire item one.

feedback sign. The response means and standard deviations are presented according to feedback sign condition in Table four, and illustrated in Figure three.

Table 4

Descriptives for Responses to Ability Study Questionnaire Item One

Sign Condition[a]	N	M (SD)
Positive (70%)	47	2.72 (0.45)
Average (50%)	44	2.23 (0.52)
Negative (30%)	50	1.70 (0.58)
Overall Mean	141	2.21 (.67)

Note. 1 = Below Average (Below 50th percentile); 2 = Average (50th percentile); 3 = Above Average (Above 50th percentile). [a] Participants not receiving performance feedback (Condition 10) are not represented in this data.

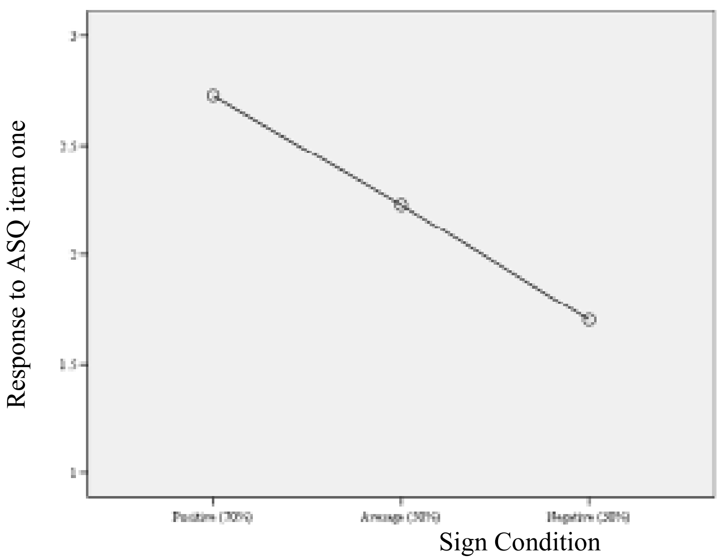


Figure 3. Plotted means for feedback sign manipulation check.

Note: 1 = Below Average (Below 50th percentile); 2 = Average (50th percentile); 3 = Above Average (Above 50th percentile). Participants not receiving feedback (Condition 10) are not represented in these data.

Next, the perceived feedback accuracy manipulation was tested by analyzing the responses to the Participant Standing Sheet item: "*We are interested in coaching the supervisor on how to provide feedback. Based on your opinion, would you say: 1) My performance feedback from the supervisor was accurate and based on the table; 2) My performance feedback from the supervisor was too low—my performance was better than she concluded; 3) My performance feedback from the supervisor was too high—my* *performance wasn't as good as she concluded*." The ANOVA results for the feedback accuracy manipulation are available in Table five. There was a main effect of feedback accuracy ($F_{(2, 91)} = 21.38$, $p < 0.01$), and no main effect for feedback sign ($F_{(2, 91)} = 1.40$, $p = 0.25$) or significant sign*accuracy interaction ($F_{(4, 91)} = 1.93$, $p = 0.12$). A Tukey post hoc test revealed that the difference between the positively distorted and accurate feedback conditions was only marginally significant (mean difference = 0.23, $p = 0.053$), while the means from all other conditions were significantly different from each other ($p < 0.01$). These means and standard deviations are presented by feedback accuracy condition in Table six, and depicted visually in Figure four.

Table 5
ANOVA Results for Feedback Accuracy Manipulation Check

	Sum of Squares	*df*	Mean Square	*F*	*p*
Intercept	330.11	1	330.11	1995.98	0.00
Sign	0.46	2	0.23	1.40	0.25
Accuracy	7.07	2	3.54	21.38	0.00
Sign*Accuracy	1.25	4	0.31	1.93	0.12
Error	15.05	91	0.17	1.89	
Total	359.00	100			

Note. Dependent variable is participant response to Participant Standing Sheet item.

Table 6
Descriptives for Responses to Participant Standing Sheet Item One

Accuracy Condition[a]	N[b]	M (SD)
Positively Distorted	34	2.12 (0.33)
Accurate	35	1.89 (0.32)
Negatively Distorted	31	1.45 (0.57)
Overall Mean	100	1.83 (0.49)

Note. 1 = My performance feedback from the supervisor was too low—my performance was better than she concluded; 2 = My performance feedback from the supervisor was accurate and based on the table; 3 = My performance feedback from the supervisor was too high—my performance wasn't as good as she concluded.
[a] Participants not receiving performance feedback (Condition 10) are not represented in this data.
[b]This item was developed after the first 45 days of data collection, therefore only 100 responses are available.

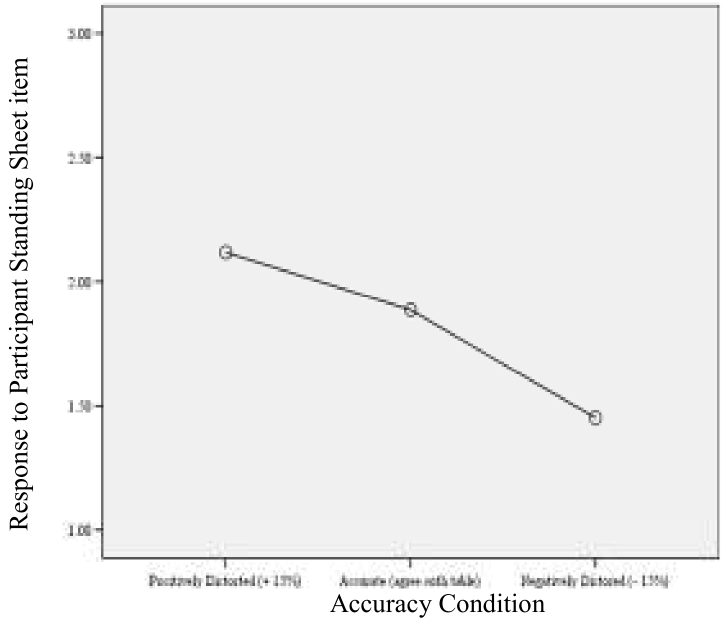


Figure 4. Plotted means for feedback accuracy manipulation check.

Note: 1 = My performance feedback from the supervisor was too low—my performance was better than she concluded; 2 = My performance feedback from the supervisor was accurate and based on the table; 3 = My performance feedback from the supervisor was too high—my performance wasn't as good as she concluded. Participants not receiving feedback (Condition 10) are not represented in this data.

Potential Confounds

Two experimental quality factors, session supervisor and the presence of the second confederate, were examined for possible influence over participant performance. Results indicated that participant performance was not affected by session supervisor ($F_{(4, 151)} = 1.41$, $p = 0.233$) or the absence of second confederate ($F_{(1, 154)} = 2.56$, $p = 0.112$).

Hypothesis Testing

The means, standard deviations and intercorrelations among study variables are presented in Table seven. Performance on the Lego task was positively related to a number of variables in the study including: gender ($r = 0.33$, $p < .01$), lego experience ($r = 0.48$, $p < .01$), lego-ability ($r = 0.46$, $p < .01$), and TSSE ($r = .40$, $p < .01$). It is important to note that when interpreting the findings reported in this paper, faster or

better performance is indicated by *larger* values due to the reciprocal transformation of time (whereas *smaller* raw completion times indicate better performance). To assess the appropriateness of the data for ANOVA and ANCOVA analyses, normality was first examined at each level of the two independent feedback variables. The data did not appear to be significantly skewed in either feedback sign
subgroups or feedback accuracy subgroups (Fisher's γ ranged from -0.07 to 0.55). Additionally it is unlikely that heterogeneity of error variance existed between groups ($F_{(8, 132)} = 0.91$, $p = 0.51$). A significant amount of performance variance was expected to occur independent of feedback condition, related to the spatial ability (e.g., Kimura, 2004) and participant experience with Legos (suspected to be higher for males). For these reasons, an attempt was made to account for pre-existing differences in ability to perform on the task.

Covariates. Participant gender and Lego-ability were considered as potential covariates in testing the hypotheses. Gender was positively related to performance on the Lego task ($r = 0.33$, $p < 0.01$), such that males were likely to exhibit faster completion times. Scatterplots depicting separate regression lines for the gender-performance relationship at different levels of the independent variables are displayed in Figures five and six. The assumption of homogeneity of regression coefficients states that the regression coefficients for each level of the independent variable(s) should be the same. A slight difference in slopes can be seen in the scatterplots, however, Glass et al (1972) suggested that researchers need only be concerned with different regression coefficients if they are significantly different from each other.

Table 7

Means and Intercorrelations

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Gender																	
2. Lego Experience	3.55	1.21	.26**														
3. Lego Index	0.98	0.17	.17*	.26**													
4. Practice Model	0.37	0.11	.31**	.37**	.74**												
5. Model 1[a]	0.25	0.08	.29**	.44**	.62**	.63**											
6. Model 2[a]	0.16	0.06	.31**	.43**	.45**	.58**	.73**										
7. Model 3[a,b]	0.12	0.04	.31*	.60**	.66**	.72**	.76**	.72**									
8. Model 3[a,c]	0.14	0.04	.32**	.43**	.32**	.54**	.65**	.71**	--								
9. Performance[a]	0.14	0.04	.33**	.48**	.46**	.64**	.76**	.94**	.91**	.90**							
10. SE Trial 1	83.95	11.47	.17*	.27**	.27**	.30**	.27**	.26**	.27	.30**	.30**						
11. SE Trial 2	87.16	11.77	.16	.23**	.23**	.21**	.25**	.24**	.37*	.22*	.27**	.66**					
12. SE Trial 3	85.65	15.39	.14	.21**	.21**	.24**	.22**	.32**	.32*	.27**	.34**	.52**	.63**				
13. SE Trial 4	83.22	18.85	.18*	.26**	.26**	.33**	.30**	.36**	.38*	.32**	.40**	.45*	.69**	.79**			
14. Overall SE	84.99	12.26	.19*	.29**	.29**	.32**	.31**	.36**	.42**	.32**	.40**	.73**	.86**	.89**	.90**		
15. Bonding Time[a]	0.33	0.14	.00	.06	.06	.05	-.05	-.03	.17	-.06	-.01	.09	-.06	.04	.02	.02	
16. Effort Avg Trials 2, 3	3.20	0.78	-.04	.09	.23**	.17	.09	.08	.30*	-.10	.06	.07	.07	.06	.12	.10	.00

Note. [a] Time has been subjected to reciprocal transformation. [b, c] Model 3 (ostrich[b]) was replaced after 44 experimental sessions with a model of similar difficulty (house[c]). * $p < 0.05$. ** $p < 0.01$.

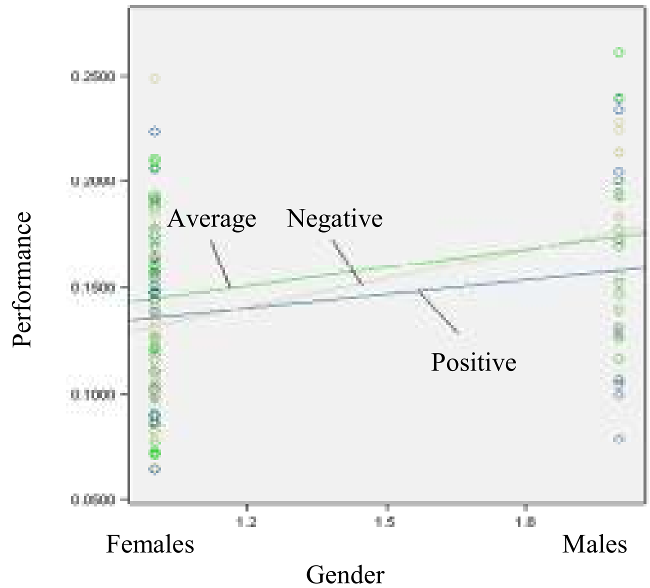


Figure 5. Separate regression lines for levels of feedback sign variable (gender).

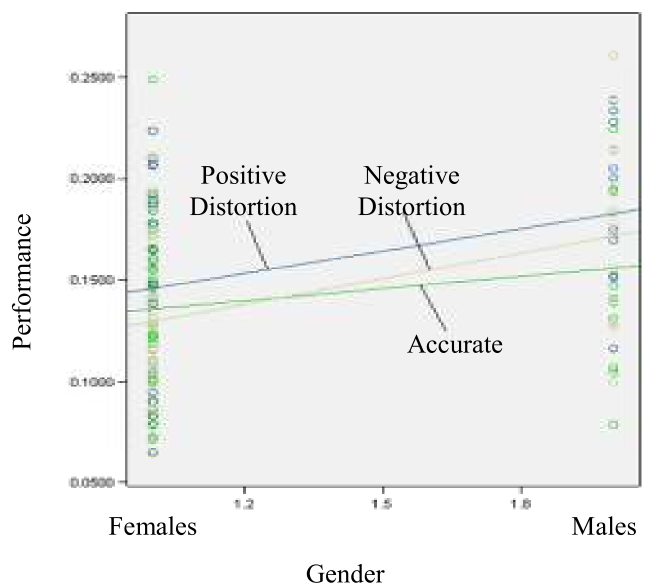


Figure 6. Separate regression lines for levels of feedback distortion variable (gender).

A Lego-ability index was created by combining the "experience with Legos or other building tasks" item from the ASQ, with the practice and trial 1 completion times, which were strongly correlated with each other ($r = 0.62$ to 0.74, $p < 0.01$). This composite was generated by first summing the transformed practice model and trial one model completion times. Next this sum was added to the “experience with Legos or other building tasks” score which had been rescaled and recoded. The Lego-ability index distribution was normal (Fisher’s $\gamma = 0.12$, standard error of skewness = 0.19), and scores on the index were positively related to the Lego performance variable ($r = 0.46$, $p < 0.01$).

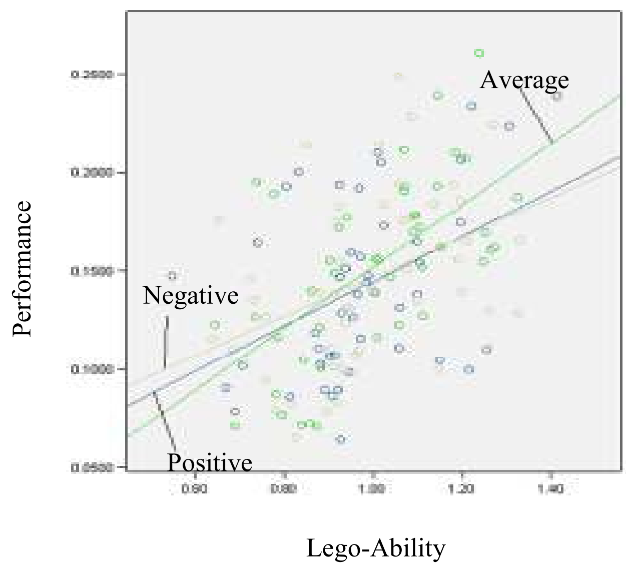


Figure 7. Separate regression lines for levels of feedback sign variable (Lego-ability).

Figures seven and eight display regression lines for the relationship between Lego-ability and performance at different levels of the independent variables. The homogeneity of regression coefficients assumption was violated by the significant difference in Lego-ability- performance slopes across the positive distortion and accurate levels of feedback accuracy ($b_{\text{positive distortion}} = .67$, $b_{\text{accurate}} = .34$, $p = 0.02$). This violation indicates that the composite slope created in the ANCOVA process may not adequately represent the data from each level of the feedback accuracy factor. Finally, multicollinearity is not expected to be operating, because Lego-ability is independent of

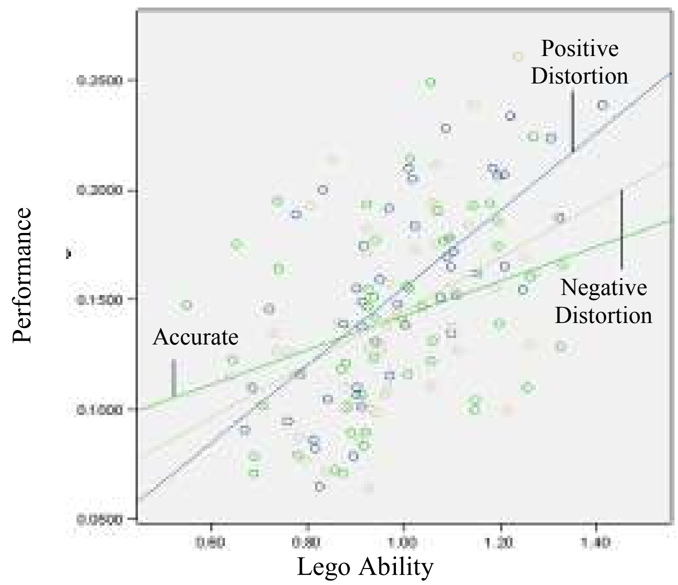


Figure 8. Separate regression lines for levels of feedback distortion variable (Lego-ability).

the independent variables, and the correlation between the two covariates is significant but weak ($r = 0.17, p < 0.05$).

The number of participants across levels of the independent variables and interaction cells was slightly unequal due to a number of unattended sessions and the aforementioned eight participants discarded due to error in procedures during the experimental sessions ($N_{IV\ level}$ ranged from 44 to 50, and N_{cell} ranged from 14 to 17). Therefore the unweighted marginal means were estimated, and Type III sums of squares were used, which has been argued as the best procedure for testing interactions in a two-way unbalanced ANOVA with no missing cells (Stewart-Oaten, 1995). Fisher's LSD post hoc tests were performed on the unweighted means, using the harmonic N to calculate the critical test value.

H1. Hypothesis one stated that the performance of participants receiving feedback would be higher than participants not receiving any feedback. The means and standard deviations for all conditions are presented in Table eight. This hypothesis was first tested with a t-test to compare performance in the control condition to average performance across all other conditions, and then the data were examined for differences between the control group and specific conditions using a one-way ANOVA. No significant difference was found between control condition performance (M = 0.131, SE= 0.011) and average performance across all other conditions (M = 0.146, SE = 0.004; $t_{(154)} = 1.27$, p = 0.21). A one-way ANOVA indicated a significant effect of condition ($F_{(9, 146)} = 2.336$, $p < 0.05$; Appendix L), and the Dunnett post-hoc test revealed a significant difference between the control condition and the condition with positive-sign and positively-distorted feedback (M = 0.170, SE = 0.011; mean difference = 0.0393, $p < 0.05$). Participants in the control condition had longer completion times than did those in the positive sign-positive distortion condition. The trends in Table eight indicate that the control condition mean was always longer than the experimental condition means, with

Table 8
Estimated Marginal Performance Means[a] and SEs by Condition and Level of Independent Variable

	Positive Distortion	Accurate	Negative Distortion	Sign Level
Positive Sign (70%)	0.170 (0.011)	0.127 (0.011)	0.126 (0.011)	0.141 (0.006)
Average Sign (50%)	0.165 (0.011)	0.136 (0.011)	0.158 (0.011)	0.153 (0.006)
Negative Sign (30%)	0.131 (0.010)	0.157 (0.010)	0.145 (0.011)	0.145 (0.006)
Distortion Level	0.155 (0.006)	0.140 (0.006)	0.143 (0.006)	
Overall Mean for Conditions 1-9	0.146 (0.006)			
Control Condition	0.131 (0.011)			

Note. [a] Means represent transformed completion times.

the exception of the positive sign-accurate and positive sign-negatively distorted conditions. To determine whether feedback for a particular level of sign or of accuracy was more effective than the control condition, a series of t-tests were conducted. First the control condition was compared to all experimental conditions combined, and then the control condition was compared to each level of the two independent variables (Table nine).

The only comparison approaching significance was that of the positively distorted conditions ($t_{(60)} = 1.79$, $p = 0.08$), which trended towards shorter completion times than the control condition. Based on the trends of longer completion times which occurred in the control condition, limited support for hypothesis one was found.

Table 9

T-tests Between Control Group and Independent Variable Levels

Control N = 15		Experimental N	t	df	p
Control vs. All groups		141	1.27	154	0.21
Control vs. Levels of Sign					
	vs. Positive Sign (70%)	47	0.82	60	0.42
	vs. Average Sign (50%)	44	1.59	57	0.12
	vs. Negative Sign (30%)	50	1.09	63	0.28
Control vs. Levels of Accuracy					
	vs. Positively Distorted	47	1.79	60	0.08
	vs. Accurate	48	0.79	61	0.44
	vs. Negatively Distorted	46	0.90	59	0.37

H2 and H3. Hypotheses two and three predicted main effects for perceived feedback accuracy and feedback sign on performance. The results of a 3 X 3 ANOVA (Table 10) did not support a main effect for accuracy ($F_{(2, 132)} = 1.69$, $p = 0.19$) or for feedback sign ($F_{(2, 132)} = 0.91$, $p = 0.41$). Next, a purer analysis was conducted, whereby participant gender and Lego-ability score were specified as covariates in a 3 X 3 ANCOVA (Table 11). This analysis also did not support a main effect for feedback accuracy ($F_{(2, 130)} = 1.84$, $p = 0.16$) or for feedback sign ($F_{(2, 130)} = 0.66$, $p = 0.52$). Thus, hypotheses two and three (including 3a) were not supported. However, support was found for Hypothesis 3b, which stated that there would be no significant difference between positive sign and average sign feedback.

Table 10
ANOVA Results for 3 X 3 Hypothesis Testing

	Sum of Squares	*df*	Mean Square	*F*	*p*	η_p^2	Observed Power
Intercept	3.00	1	3.003	1655.17	0.00		1.00
Sign	0.00	2	0.002	0.91	0.41		0.20
Accuracy	0.01	2	0.003	1.69	0.19		0.35
Sign*Accuracy	0.03	4	0.007	3.72	0.01	.101	0.88
Error	0.24	132	0.002				
Total	3.278	141					

Note. Dependent variable is performance.

Table 11
ANCOVA Results for 3 X 3 Hypothesis Testing

	Sum of Squares	*df*	Mean Square	*F*	*p*	Observed Power
Intercept	0.001	1	0.001	0.75	0.39	0.14
Gender	0.016	1	0.016	11.56	0.00	0.92
Lego-ability	0.037	1	0.037	27.04	0.00	0.99
Sign	0.002	2	0.001	0.66	0.52	0.16
Accuracy	0.005	2	0.002	1.84	0.16	0.38
Sign*Accuracy	0.012	4	0.003	2.13	0.08	0.62
Error	0.176	130	0.001			
Total	3.278	141				

Note. Dependent variable is performance.

H4. The interaction between feedback accuracy and sign was found to be significant in the ANOVA results ($F_{(4,132)} = 3.72$, $p < 0.01$). Figure nine plots the significant interaction between accuracy and sign. Hypothesis 4a specified that positively distorted-positive feedback would result in lower performance than positively distorted-negative feedback, and a post-hoc planned comparison indicated a significant difference between these two conditions (Fisher's LSD; $t_{crit\,(132,\,.05)} = 0.027$, difference in means = .039). However, the interaction occurred in the opposite direction of hypothesis 4a, as positively distorted-positive feedback resulted in noticeably better performance than positively distorted-negative feedback. It is also interesting to note in the plotted means the apparent superiority of performance following positively distorted feedback compared to other accuracy/distortion conditions in the positive sign (70%) and average sign (50%) levels vs. the relatively low performance following positively distorted feedback compared to other conditions in the negative sign (30%) condition.

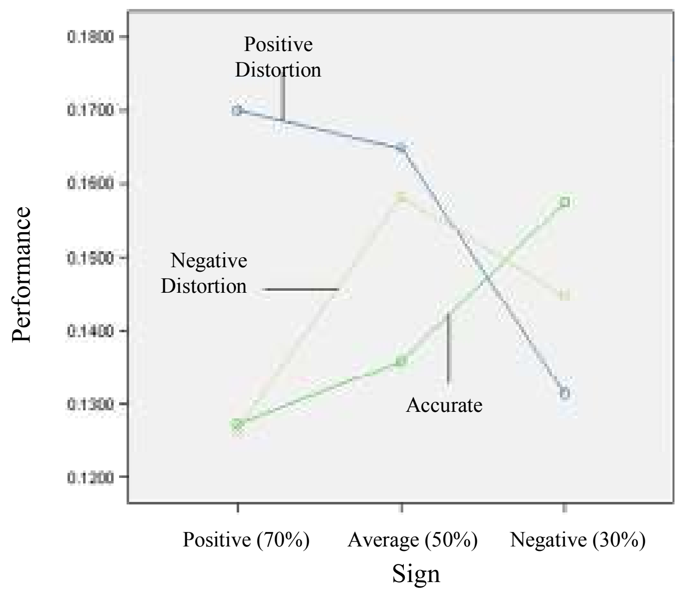


Figure 9. ANOVA interaction between feedback sign and feedback accuracy.

Hypothesis 4b stated that negatively distorted-negative sign feedback would result in lower performance than negatively distorted-positive or average feedback. Post-hoc planned comparisons did not indicate a significant difference in performance between these conditions (Fisher's LSD; $t_{crit\ (132,\ .05)}$ = 0.027, difference in means= 0.013, 0.019), failing to provide support for this hypothesis. Inspection of trends suggested that while the negatively distorted feedback tended to yield slightly higher performance in the average sign condition than in the negative sign condition, it also appeared to result in the lowest performance in the positive sign condition (in conflict with direction of Hypothesis 4b).

Finally, hypothesis 4c stated that when feedback is perceived to be accurate, positive sign feedback would be associated with significantly higher performance than either average feedback or negative feedback. Feedback perceived as accurate was also predicted to result in significantly higher performance in the average sign condition relative to the negative sign condition. Post hoc planned comparisons indicated that there was a significant difference between the accurate positive and accurate negative feedback conditions (Fisher's LSD; $t_{crit\,(132,\,.05)} = 0.027$, difference in means $= 0.030$), but this interaction occurred in the opposite direction of the hypothesis, such that accurate negative feedback resulted in significantly higher performance than accurate positive feedback. No difference existed between the accurate-positive and accurate-average conditions (difference in means $= 0.011$) or between the accurate-average and accurate-negative conditions (difference in means = 0.021).

Interestingly, when gender and Lego-ability score were entered as covariates, the *perceived accuracy**sign interaction fell short of significance ($F_{(4,\,130)} = 2.13$, $p = 0.08$; means available in Appendix L). Comparison of the means (Figure 10) suggests that variability in performance was slightly reduced, as means in all negative sign conditions were higher, and performance in the accurate feedback conditions seemed to be less influenced by level of feedback sign. Despite these shifts, the performance trends across conditions were relatively stable, as positively distorted feedback remained as the most effective feedback across the positive sign (70%) conditions and the least effective feedback in the negative sign (30%) conditions.

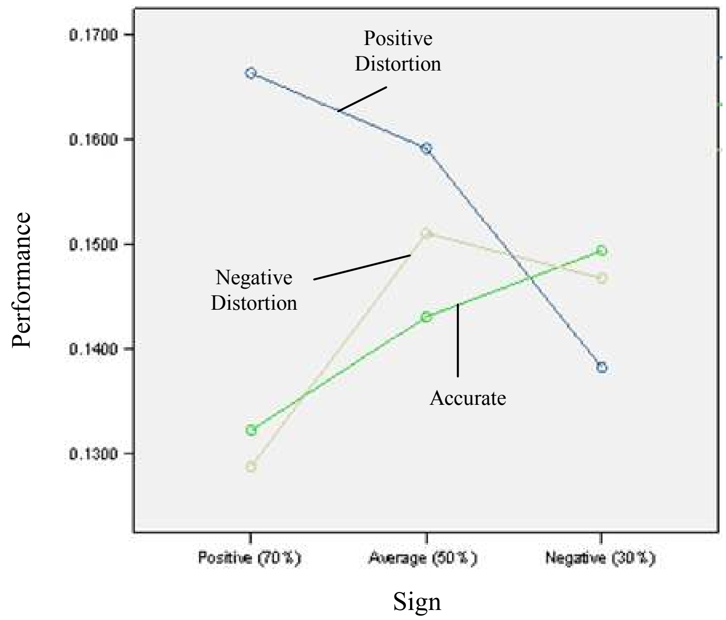


Figure 10. ANCOVA interaction between feedback sign and feedback accuracy.

*Exploratory***.** Overall task-specific self efficacy was proposed as a mediator between type of feedback and performance, and as an outcome affected by the type of feedback. Hierarchical mediated regression was used to test this exploratory hypothesis according to the procedure outlined by Baron and Kenny (1986). In order to test the relationship between the independent variables and the dependent variable, the levels of feedback sign and feedback accuracy were dummy coded into four new variables (Table 12). Table 13 presents the correlations between the exploratory hypothesis variables.

Table 12
Dummy Coding of Independent Variables

Sign IV	Dummy 1	Dummy 2
Positive	1	0
Accurate	0	1
Negative	0	0

Accuracy IV	Dummy 3	Dummy 4
Positive	1	0
Accurate	0	1
Negative	0	0

Table 13
Correlations Between Performance and Dummy-Coded Variables

	1	2	3	4	5	6	7
1. Dummy 1							
2. Dummy 2	-0.48**						
3. Dummy 3	0.01	-0.02					
4. Dummy 4	-0.03	0.03	-0.51**				
5. Performance[a]	-0.07	0.10	0.14	-0.08			
6. Overall TSSE	0.08	-0.01	0.06	-0.04	0.40**		
7. Lego-ability	-0.04	0.07	0.05	-0.03	0.46**	0.16*	
8. Gender	-0.01	-0.02	-0.05	-0.02	0.33**	0.19*	0.17*

Note. [a] Time has been subjected to reciprocal transformation. * $p < 0.05$. ** $p < 0.01$.

The results of the mediated hierarchical regression are in Table 14. First, the relationship between TSSE (mediator) and performance (criteria) was tested and found to be significant ($R^2 = 0.16$, $p < 0.01$). Next, feedback accuracy and feedback sign (predictors) were analyzed as predictors of TSSE (mediator). This relationship could not be established ($R^2 = 0.01$, $p = 0.85$). The independent variables were then entered along

Table 14
Results of Mediated Hierarchical Regression Analyses

	Z predict Y	X predict Z	Mediation	Full vs. partial	Full vs. partial
Sign					
Dummy 1		0.09	-0.06	-0.06	
Dummy 2		0.03	0.07	0.07	
Accuracy					
Dummy 1		0.05	0.12		0.12
Dummy 2		-0.01	-0.02		-0.02
TSSE	0.40^{**}		0.40^{**}	0.41^{**}	0.40^{**}
R^2	0.16^{**}	0.01	0.19^{**}	0.18^{**}	0.18^{**}
F	0.01	0.34	6.43	9.73	10.07

Note. Dependent variable is performance. The independent variables are notated as X, dependent variable as Y, and mediator as Z. $^{**}p < 0.01$.

with the mediator, as predictors of performance, but mediation was not supported (*R*-square for *perceived accuracy* = 0.02, $p = 0.19$; sign = 0.01, $p < 0.56$; and TSSE = 0.16, $p < 0.01$). Because this hypothesis was exploratory, tests for partial vs. full mediation were also conducted, but did not yield any significant findings.

The second exploratory hypothesis predicted significant effects of perceived feedback accuracy on TSSE. To analyze this possibility, first a t-test compared TSSE in the control condition to average TSSE across all other conditions, and then the data were examined for differences between the control group and specific conditions using a one-way ANOVA. No significant differences in TSSE means were found between the control and combined feedback conditions (($t_{(154)} = 1.37$, p = 0.14), or between control condition and specific feedback conditions ($F_{(9, 145)} = 0.42$, $p = 0.924$). Next, a 3 X 3 ANOVA was conducted to test for significant interactions between feedback sign and feedback accuracy. Again, the analysis did not yield any significant main effect for feedback sign (F(2, 131) = 0.42, p = 0.66), perceived feedback accuracy ($F_{(2, 131)} = 0.30$, $p = 0.74$), or for a sign*accuracy interaction ($F_{(4,131)} = 0.48$, $p = 0.75$; Appendix L). Despite insignificance, it is interesting to examine the trends from this analysis (Figure 11), specifically noting the larger differences between levels of *perceived accuracy* when feedback sign is average. The association of accurate feedback with the lowest TSSE in the positive and average sign conditions was unexpected.

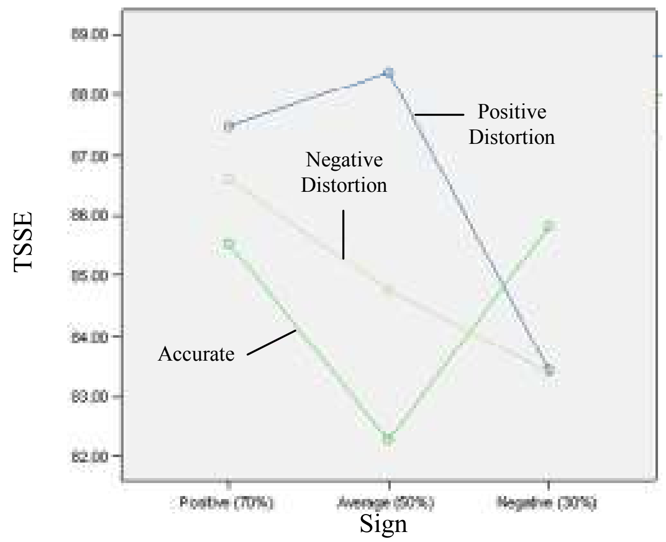


Figure 11. Insignificant sign-accuracy interaction on overall TSSE.

Finally, in the negative sign condition, the similar effects of positively and negatively distorted feedback relative to feedback perceived as accurate are somewhat puzzling and will be addressed further in the discussion.

Follow-Up Analyses

As detailed in the method section, measures of self-assessment and effort expended were collected after every performance trial. The relevance of the effort expended rating to the *desired response* and *intended response* constructs in the Ilgen, et al (1979) and Kinicki, et al (2004) model sparked interest in examining these variables, specifically in the trials before or after the performance feedback was provided.

Therefore several follow-up analyses were conducted on an exploratory basis in order to expose more qualitative information about the feedback processes observed in this experiment. The variables representing processes before the performance feedback was provided included the self-report of effort expended on model one, and SA of model one. Following the feedback administration, the follow-up variables of interest included TSSE measured in anticipation of the next model, and the effort expended on model two.

Considering the potential role of SA in the recipient's perception of accuracy (see *Framing Perceived accuracy*), it is reasonable to assess the extent to which a person's SA differs from the condition-defined levels of feedback which were provided on the same trial. Therefore, each participant's SA response (percentile at which the participant ranked his or her performance relative to other participants who had completed the task) gathered immediately after the first official trial but before the provision of performance feedback, was assessed in comparison to the percentile indicated by the feedback as specified by the conditions. The condition-specified percentile against which the SAs were compared took both sign and distortion manipulations into account (e.g., negative distortion- positive sign condition rated the participant at 70-15 = 55th percentile).

Positive discrepancies between trial one SA and the feedback provided after trial one indicated that the participant's SA was higher than the condition-defined feedback which was provided for the same model (e.g., Participant rated himself or herself as being in the 60th percentile, and the feedback provided for the particular condition indicated that he or she was in the 50th percentile). The correlations between the follow-up variables are presented in Table 15.

Table 15
Correlations Between Overall Performance and Follow-Up Variables

	1	2	3
1. Performance[a]			
2. TSSE Trial 2	0.27*		
3. Effort Trial 2	-0.03	-0.13	
4. Positive SA Trial 1-Feedback Discrepancy[b]	0.05	0.17**	0.08
5. Negative SA Trial 1-Feedback Discrepancy[c]	-0.07	0.30	0.23

Note. [a] Time has been subjected to reciprocal transformation. [b] N = 109 [c] N = 38 * $p < 0.05$ ** $p = 0.07$

The magnitude of these positive SA – feedback discrepancies was positively related to self-reported TSSE, although the significance was marginal ($r = 0.17$, $p = 0.07$). This trend indicates that the lower the feedback was relative to the participants' self-ratings (i.e., the participants should have interpreted this as overrating themselves), the higher their TSSE was for the upcoming model. When the feedback was only slightly below the participants' SA ratings, the participants' TSSE on the subsequent trial was lower.

Negative trial one SA-feedback discrepancies represent cases in which the participant's SA was lower than the percentile indicated by the feedback condition (e.g. Participant SA indicated 60^{th} percentile, and feedback from supervisor concluded that he or she had performed in the 70^{th} percentile). The correlational analyses suffered from a low N in this condition, because there were fewer cases when the participant rated themselves lower than the provided feedback (consistent with the SA literature). Despite the lack of power, a sizeable correlation was found between the magnitude of negative discrepancies and ratings of TSSE before the subsequent trial ($r = 0.30$, $p = 0.07$). Specifically, the smaller the discrepancy (i.e., the less far above the SA the feedback was) the higher their TSSE ratings were on the subsequent model, and the higher their reported effort after the second trial (r = 0.23, p = 0.16).

In the interest of bringing additional meaning to these relationships the next step was to determine whether the feedback sign or accuracy manipulations had affected the trial two follow up variables. A 3 X 3 ANOVA with trial two TSSE did not reveal a main effect for feedback sign ($F_{(2, 131)}$ = 0.59, p = 0.56), feedback accuracy ($F_{(2, 131)}$ = 0.42, p = 0.66), or interaction ($F_{(4, 131)}$ = 0.89, p = 0.89; Appendix L). The trends across condition for trial two TSSE were similar to those previous reported for overall TSSE.

Next, the trial two effort ratings were assessed by a 3 X 3 ANOVA to identify whether the participants intent to improve had been affected by the feedback conditions. No significant findings resulted from this analysis either for feedback sign ($F_{(2, 132)}$ = 0.89, p = 0.41), perceived feedback accuracy ($F_{(2, 132)}$ = 1.03, p = 0.36), or interaction ($F_{(4, 132)}$ = 0.68, p = 0.68; Appendix L).

Discussion

The principal concern with performance feedback is how it may best serve to enhance performance. The feedback model (Ilgen, et al., 1979; Kinicki, et al, 2004) implicates *source credibility* and the *feedback-rich environment* as critical elements in the process explaining recipient accuracy perceptions and responses. Building on this theoretical framework, the present study examined different levels of both perceived feedback accuracy and feedback sign. The results suggest that the highest performance occurred in the positive distortion-positive sign feedback condition, although significant differences were observed between the positively distorted-positive sign feedback condition and the positively distorted-negative sign feedback and control conditions. Although control condition performance was lower than performance in almost all other conditions (tying with the low positively distorted-negative sign condition performance), the scarcity of significant differences between the control condition and other feedback conditions suggests that sometimes scenarios with no feedback may yield similar performance to scenarios with performance feedback (this issue is further addressed in the study limitations). The tendency of positively-distorted feedback to yield the highest observed performance in the average and positive conditions generated interest in a potential significant performance difference between these conditions (i.e., perceived positive distortion of non-negative feedback) and all other feedback scenarios created in the study. A post-hoc test indicated that, indeed, performance in the perceived positively distorted-average and positive sign conditions (M = .1676 or 5m58s) was significantly higher ($t_{(154)} = 3.28$, $p < 0.01$) than performance in all other feedback conditions (M = 0.1389 or 7m11s). The following discussion argues that this perhaps explains why supervisors tend to positively distort both sign and accuracy.

In the literature a positive relationship has been demonstrated between *perceived accuracy* and subsequent process variables such as *desire to respond* and performance (Kinicki, et al, 2004). This positive relationship was not supported in the present experiment, as indicated by the lack of a main effect for *perceived accuracy* on any dependent variable. Rather, performance differences were observed when *perceived accuracy* interacted with feedback sign. The trends in the average-sign condition are particularly telling since, by definition, average performance is more commonly encountered in the real world than high or low performance. In these average sign conditions, performance tended to be higher when feedback was perceived as positively or negatively distorted compared to feedback perceived as accurate. Overall, perceived positive distortion in the positive sign condition benefited performance, whereas in the negative sign condition perceived positive distortion resulted in significantly lower performance. Perceived negative distortion yielded slightly lower performance when crossed with the positive sign condition compared to the negative sign condition, but the effect was not significant.

Taking perceived positive and negative distortion into account, the trends for these two sides of perceived inaccuracy suggest that perceived inaccuracies were actually beneficial when they exaggerated performance in the direction of the feedback sign (e.g., positive distortion of positive sign feedback, and negative distortion of negative sign feedback). Distortions negatively affected performance when the distortion direction was contrary to the basic feedback sign (e.g., negative distortion of positive sign feedback). These findings are fascinating: As long as the distortion is generally representing the level of performance (e.g., positive distortion of non-negative feedback, negative distortion of negative sign) recipients are more likely to respond, perhaps because they feel that they are receiving what they deserve. Recipients seem to respond poorly when the feedback does not seem warranted (e.g., negatively distorted when sign is positive or positively distorted when it is clear that the sign is generally negative). Distortions which are contradictory to the level of performance may reduce the *desire to respond* in the positive sign condition because the recipient suspects that it is pointless to perform at a high level if the performance will be negatively distorted. Recipients in the perceived positively distorted-negative sign condition may also be less motivated to respond because they anticipate that the positive distortion will hide their low performance.

Results from the conditions in which feedback was perceived to be accurate were also surprising. Specifically, *perceived accuracy* of feedback led to higher performance in the negative sign condition, followed by lower performance in the average condition, and lowest performance in the positive sign condition. This unexpected finding contradicts previous research in which negative feedback has resulted in low subsequent performance and positive feedback yielded high subsequent performance. However, the previous research has never attempted to ensure that the feedback is perceived as being accurate. For example, negative feedback is often dismissed or rejected (Kluger & DeNisi, 1996), which likely indicates that the recipient's perception of accuracy was low. The present findings suggest that if perceived accuracy is strong, negative feedback may actually be beneficial for performance. According to the feedback process model, *perceived accuracy* operates through the *desire to respond* and *intended response* to affect performance. In the accurate-negative sign condition the feedback is perceived as accurate and the negative sign indicates that a response is needed, which is likely to trigger a stronger *desire to respond*. On the other hand the accurate-positive sign condition met the model's *perceived accuracy* qualification, but it may not have signaled the need to respond since the feedback indicated performance was good. This possibility has implications for revision of the model's *feedback rich environment* antecedent.

The *feedback-rich environment* antecedent posited by the feedback process model emphasizes specific, frequent, and positive feedback. The levels of perceived sign and objective sign represent important components which are overlooked in the model, and as evidenced by the interactions just discussed, positive feedback does not necessarily lead to better performance. The feedback model would also benefit from including as an antecedent to *perceived accuracy* the discrepancies between self vs. external evaluations of performance. Specifically, the magnitude and sign of these discrepancies resulted in differing relationships. The discrepancies between participant self-evaluation of trial one performance and the bogus feedback provided on the trial one performance should have been interpreted by the participants as overrating or underrating their performance (this has no meaning in an absolute sense because the feedback was fictional, but the participants were not aware of that). Considering this, participants who "overrated" themselves were significantly more likely to report higher TSSE on the subsequent model than did participants who's SA was closer to the feedback. It is possible that larger discrepancies triggered a stronger *desire to respond*, and higher levels of TSSE indicated goals and/or dedication to overcome the discrepancy. Another possible explanation is that larger discrepancies were perceived as inaccurate feedback. Low *perceived accuracy* may have rendered the feedback easy to dismiss, while indicating high TSSE served to self-validate the recipient. By contrast, the smaller overrating discrepancies might have been difficult to dismiss and therefore perceived as an accurate portrayal of the recipient's performance, causing reduced TSSE on the subsequent trial.

Discrepancies in which the recipient had highly underrated himself or herself were associated with lower TSSE and lower effort expended on the subsequent trial relative to participants who's SAs were closer to the feedback they received. This may represent a motivational quality of feedback which reinforces one's self-appraisal, but tt is not clear why receiving feedback that is higher than expected would be associated with lower TSSE and lower effort.

The variance accounted for by the Lego-ability index covariate is difficult to interpret. Random assignment should have dispersed participant Lego-ability somewhat equally across the conditions. The strong significance found for Lego-ability as a covariate, and its reduction of the interaction significance can be interpreted in two ways. First, the interaction may have been due in part to an uneven spread of Lego-ability across conditions. Analyses of the two components of the Lego-ability index did not indicate a significant difference in transformed average practice model and model 1 completion times ($F_{(8, 132)} = 1.80$, $p = 0.08$) or in self-reported experience with Legos ($F_{(8, 132)} = 0.80$, $p = 0.61$) across conditions. A second interpretation of the covariate effect on significance is that accounting for the individual differences in Lego-ability (including practice model and trial one model performance) may have removed variance attributable to some motivational aspect of the participants' previous performances in addition to their general Lego-ability. In light of these findings, researchers are cautioned against using previous performance as a control or covariate in investigations of feedback and recipient motivation to respond.

It would perhaps be useful to examine the repeated measure data using a hierarchical conceptualization. For example, multiple data points were collected for each recipient. It is likely that different levels of *perceived accuracy* and perceived sign affect the way in which performance, TSSE, and effort expended vary over time. Additionally, it would be interesting to assess how accurate participants were in estimating ability on the TSSE by comparing their scores with their performance percentile rank in the study sample. These additional findings might facilitate interpretation of the role of TSSE in the feedback process, which was not very clear in the analyses (significantly related to performance, but not significantly related to feedback perceptions as a mediator or dependent variable).

Implications

The results of this study serve to clarify aspects of the widely-held assumption that an overarching positive relationship exists between the accuracy of performance appraisal and subsequent recipient improvement and of the inescapable dysfunctional consequences of misleading, incomplete or inaccurate feedback (London, 2003). A number of studies have demonstrated the prevalence of rating distortion in PA, especially positive distortion or inflation (e.g., Benedict & Levine, 1988; Fisher, 1979; Guendfeld & Weissenberg, 1966; Tesser & Rosen, 1971; Waung & Highhouse, 1997). This is often construed as an unfavorable rating outcome, because employees won't be informed of the quality of their true performance. Instead of holding raters accountable for the absolute or objective accuracy of their ratings, it is likely that recipient responses to feedback are equally or more likely to be based on underlying cognitive processes and perceptions, than on objective feedback characteristics (Kinicki, et al 2004). The present study demonstrates that perceptions of positive distortion are actually beneficial for recipient responses to non-negative feedback. Regarding negative feedback, perceptions of accuracy do appear to be related to better performance, although more research is needed to substantiate this trend. Following this, it appears that equal prioritization of perceived feedback accuracy and absolute PA rating accuracy in research will be necessary to identify the performance management strategy which is most beneficial for the employees and organization (e.g., perhaps perceived positive distortion is beneficial for future performance in terms of recipient cognitive response and performance processes, while absolute PA rating accuracy is important for evaluation of organizational objectives).

Study Strengths and Limitations

There are several strengths and limitations to note in the present research. This study improved upon previous research by avoiding dependency on self-report data, and using a more complex measure of performance. The use of completion time and errors captures more aspects of performance than a single Likert-type self-rating of performance, time-estimation deviations or math test number of correct. The experimental conditions represented the first attempt to manipulate perceived feedback accuracy, and together with manipulations of perceived sign the experimental design allowed for a more indepth picture of sub-surface processes, and the use of an experiment provided for the interpretation of causal trends instead of simple associations.

The first limitation of this study is that the positively distorted feedback vs. accurate feedback manipulation was not as strong as other levels of the independent variables. One way to improve the strength of the positive distortion perception could be to change the introduction by a second confederate to avoid building up the supervisor's credibility, in addition to the distorted feedback comments. It is also possible that because supervisors tend to positively distort PA ratings (Benedict & Levine, 1988), it is difficult for recipients to perceive positive distortion as anything but accurate.

Although the data generally support the effectiveness of the independent variable manipulations, it is not possible to completely rule out the impact of other variables covarying with the independent variables of interest. The experimental conditions may have activated other cognitive processes or perceptions. For example, the feedback manipulation of positive distortion could affect performance by influencing the supervisor's likeability (e.g., participant's feelings towards supervisor contribute to higher performance) or encouraging a self-fulfilling prophecy (e.g., supervisor communicates that she thinks the participant's performance was higher, participant subsequently performs at a higher level). The data collected do not allow analysis of these alternate explanations, but they are worthy of consideration in the future development of feedback perception and process theory.

A second limitation is that while feedback specificity was purposely manipulated (high in perceived accurate, low in perceived distortion conditions), the feedback was not designed to be developmental for the participants (e.g., providing suggestions or strategy for improvement). Developmental quality is likely to be an important component of the feedback process, perhaps as an aspect of the *feedback-rich environment* which may moderate the relationship between *perceived accuracy* and *desire to respond.* Third,the participants were never reminded of the gift certificate incentive beyond the introductory task instructions. Kinicki et al (2004) note that the recipient's interest in outcomes will impact *the desire to respond* and *intended response.* However, Cognitive Evaluation Theory (CET; Deci, 1971) argues that reward systems exert negative influences on intrinsic motivation—reducing performance in the long-term (Lepper, Green & Nisbet, 1973) unless the rewards were designed to increase the employees' perceptions of self-confidence and self-determination. The incentive may have therefore functioned appropriately in this case. The fourth limitation is based on an insignificant correlation between effort expended (follow-up variable) and performance. Higher self-report of effort expended was not associated with higher performance (r = -0.04, p = 0.62). The insignificant relationship with self-reported effort indicates that attempts to perform at a higher level did not necessarily result in higher performance. It is possible that the Lego task or method of performance measurement may not have been conducive to performance improvement attempts. A second explanation may be that self-reported effort, absent an improved capacity to perform, may not have been important in improving performance. Finally, the study variables should be examined in the context of a task in which initial performance across participants is slightly less variable. The Lego task was selected because it provided a performance metric that represented two aspects of performance (time and error), it was likely to be perceived as interesting by the participants, successful performance required both cognitive ability and psychomotor skills, and it did not require any training. However, task characteristics, performance setting (lab vs. field) and occupation type are likely to interact with the general usefulness of feedback as well as its perceived accuracy (e.g., the speed component of the task, short time-frame of the experimental session and relatively consequence-free laboratory). Additionally, even routine assembly tasks in the work place, represented here by a Lego task, are likely to involve more dimensions of performance than merely speed and error. These task issues may limit the generalizability of study findings, but they also present interesting directions for future research of perceived feedback accuracy.

Future Directions

As suggested by the results of this study, the performance feedback domain is still in need of a great deal of theoretical development. The complex results presented here serve as an indicator of the intricate cognitive processes operating in the feedback process. A high priority in this area of research is continued expansion and refinement of the feedback process model (Ilgen, et al, 1979). In particular, attention should be given to four main areas: 1) The role of feedback sign and developmental quality in the *feedback-rich environment*; 2) The role of self-other feedback discrepancies in defining a recipient's *perception of feedback accuracy*; 3) The role of the recipient's capacity for task performance , perhaps between *intended response* and *desire to respond*; and 4) The identification of additional cognitive factors which interact with perceptions of feedback, such as mood, and personality. Recipient mood can be affected by feedback (e.g., Goldstein & Strube, 1994), and it can influence the level of performance after feedback (e.g., Anshel, 1988). Mood has been shown to lower the *perceived accuracy* of feedback when the consistency was low between recipient mood state and the favorability of feedback (Hammer & Stone-Romero, 1996). Personality factors such as negative affectivity and narcissism can also create differences in perceptions of feedback accuracy. For example, persons high on Narcissism are likely to accept and take credit for feedback of success, but attribute failure feedback to external factors (Stucke, 2003) which is likely to lower the *desire to respond*. Conscientious recipients are likely to have a higher *desire to respond* as indicated by demonstrated motivation to learn and improve after feedback (Colquitt & Simmering, 1998). Future work should also be done to examine the role of self-esteem, as high self-esteem is sometimes linked to low motivation to persist after failure (e.g., Baumeister & Tice, 1985), in addition to being linked to higher persistence after failure (e.g., Di Paula & Campbell, 2002). Of particular relevance to *perceived accuracy*, self-esteem has been shown to interact with the psychological prominence of a person's strengths and weaknesses, such that negative feedback can cause low-self-esteem persons to focus on their weaknesses (Dodgson & Wood, 1998). This type of response is likely to increase acceptance of the negative feedback, but perhaps not improve performance.

A number of demographic variables have been studied in the PA literature, and they warrant application in feedback theory as well. Perceived biases related to feedback source-recipient gender and race have been shown to affect perceptions of accuracy in different ways (e.g., Geddes & Konrad, 2003; Britt & Crandall, 2000). For example, males tend, more so than women, to self-handicap in response to performance feedback when the task is allegedly reflective of the participant's ability (Dietrich, 1995). Johnson and Helgeson (2002) investigated gender differences and reported that men's self-esteem was unaffected by feedback sign, whereas women's self-esteem slightly improved after positive feedback and was significantly lower after receiving negative feedback. These findings are likely related to differences in tendencies to accept or reject feedback based on *perceived accuracy*. Finally, extension of the present study is necessary to: 1) determine the extent to which *perceived accuracy* can facilitate the use of negative feedback to improve performance, and respond to the call for the development of a method which facilitates the effectiveness of feedback for poor performance (Goodman & Wood, 2004), and 2) examine experimental feedback manipulations across tasks which represent varying occupations.

The integration of PA and feedback research objectives is warranted in order to begin filling in the picture of the entire performance appraisal process. Research which isolates only one side of performance management (i.e., only PA or only feedback processes) fails to capture the entire progression from rater assessment of performance to recipient performance, which restricts theoretical development. Two examples illustrating this issue are these: 1) Perhaps perceived feedback accuracy may be affected by rater attempts to delay presentation of their ratings (aspect of PA rating process), with consequence to subsequent performance; and 2) Recipients may perceive feedback accuracy differently and subsequently *desire to respond* differently, dependent on the rater's suspected motive for distorting (e.g., avoidance of conflict, misremembering subordinate's performance two months prior). These examples demonstrate that research of the PA process alone (e.g., rating inflation and training) cannot determine the consequences of the PA ratings, unless dissemination of performance feedback and subsequent performance are also integrated. Investigation of the relationship between absolute accuracy of supervisor ratings and the perceived feedback accuracy of the recipient will also be critical to capture and model. Furthermore, it is important to realize that the rating of performance and the perception of/response to feedback are process constructs imbedded in a larger organizational context of antecedents and performance outcomes. In support of these objectives, future research should manipulate perceived feedback accuracy in an organizational context, in which a recipient's performance is linked to real benefits and consequences.

References

Albright, M.D., & Levy, P.E. (1995). The effects of source and performance rating discrepancy on reactions to multiple raters. *Journal of Applied Social Psychology*, 25(7), 577-600.

Alvarez, A.I.G & Bernal, A.O. (1999). Work Feedback and the accuracy of performance self-assessments. *Psychology in Spain*, 3(1), 126-137.

Alvero, A.M., Bucklin, B.R.,& Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings. *Journal of Organizational Behavior Management*, 21(1), 3-29.

Anshel, M.H. (1988). The effect of mood and pleasant versus unpleasant information feedback on performance motor skill. *Journal of General Psychology*, 115(2), 117-129.

Ashford, S.J. (1989). Self-assessments in organizations: A literature review and integrative model. *Research in Organizational Behavior*, 11, 133-174.

Atwater, L.E., Rousch, P., & Fischthal, A. (1995). The influence of upward feedback on self and follower ratings of leadership. *Personnel Psychology,* 48, 34-60.

Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.

Bandura, A. (1980). Gauging the relationship between self-efficacy judgment and action. *Cognitive Therapy and Research*, 4(2), 263-268.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*, 25, 729-735.

Bannister, B.D. (1986). Performance outcome feedback and attributional feedback: Interactive effects on recipient responses. *Journal of Applied Psychology*, 71(2), 203-210.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. Journal of Personality and Social Psychology, 51, 1173-1182.

Baumeister, R.F., & Tice, D.M. (1985). Self-esteem and responses to success and failure: Subsequent performance and intrinsic motivation. *Journal of Personality*, 53, 450-467.

Benedict, M.E., & Levine, E.L. (1988). Delay and Distortion: Tacit influences on performance appraisal effectiveness. *Journal of Applied Psychology*, 73, 507-514.

Blakely, G. L. (1993). The effects of performance rating discrepancies on supervisors and subordinates. *Organizational Behavior and Human Decision Processes, 54*, 57-80.

Brett, J.F., & Atwater, L.E. (2001). 360 degree feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, 86, 930-942.

Britt, T.W., & Crandall, C.S. (2000). Acceptance of feedback by the stigmatized and non-stigmatized: The mediating role of the motive of the evaluator. *Group Processes & Intergroup Relations*, 3(1), 79-95.

Brosvic, G.M., & Finizio, S. (1995). Inaccurate feedback and performance on the Mueller-Lyer Illusion. *Perceptual & Motor Skills*, 80(3), 896-898.

Burke, R.J., Weitzel, W., & Weir, T. (1978). Characteristics of effective employee performance review and development interviews: Replication and extension. *Personnel Psychology*, 31, 903-920.

Cannon-Bowers, J.A. (1988). Investigation of the construct validity of self-efficacy: Conceptualization, measurement, and causal properties. Unpublished doctoral dissertation, University of South Florida, Tampa.

Chen, G., Gully, S.M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, 4, 62-83

Colquitt, J.A., & Simmering, M.J. (1998). Conscientiousness, goal orientation, and motivation to learn during the learning process: A longitudinal study. *Journal of Applied Psychology*, 83(4), 654-665.

Cummings, L.L., Schwab, D.P., & Rosen, M. (1971). Performance and knowledge of results as determinants of goal setting. *Journal of Applied Psychology*, 56(6), 526-530.

Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

DeNisi, A. S. & Kluger, A.N. (2000). Feedback effectiveness: Can 360-degree appraisals be improved. *Academy of Management Executive*, 14(1), 129-139.

Dietrich, D. (1995). Gender differences in self-handicapping: Regardless of academic or social competence implications. *Social Behavior & Personality*, 23(4), 430-410.

Di Paula, A., & Campbell, J.D. (2002). Self-esteem and persistence in the face of failure. *Journal of Personality & Social Psychology*, 83(3), 711-724.

Dodgson, P.G., & Wood, J.V. (1998). Self-esteem and the cognitive accessibility of strengths and weaknesses after failure. *Journal of Personality &Social Psychology*, 75(1), 178-197.

Fedor, D.B., Davis, W.D., Maslyn, J.M, & Mathieson, K. (2001). Performance improvement efforts in response to negative feedback: The roles of source power and recipient self-esteem. *Journal of Management*, 27, 79-97.

Fisher, C.D. (1979). Transmission of positive and negative feedback to subordinates: A laboratory investigation. *Journal of Applied Psychology*, 64, 533-540.

Geddes, D., & Konrad, A.M. (2003). Demographic differences and reactions to performance feedback. *Human Relations*, 56(12), 1485-1513.

Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research,* 42, 237-288.

Goldstein, M.D., & Strube, M.J. (1994). Independence revisited: The relation between positive and negative affect in a naturalistic setting. *Personality & Social Psychology Bulletin*, 20(1), 57-64.

Goodman, J.S. & Wood, R.E. (2004). Feedback specificity, learning opportunities, and learning. *Journal of Applied* Psychology, 89(5), 809-821.

Greller, M.M. & Parsons, C.K. (1992). Feedback and feedback inconsistency as sources of strain and self-evaluation. *Human Relations*, 45(6), 601-621.

Gruendfeld, L.W., & Weissenberg, P. (1966). Supervisory characteristics and attitudes towards performance appraisals. *Personnel Psychology*, 19, 143-151.

Hillman, L.W., Schwandt, D.R., & Bartz, D.E. (1990). Enhancing staff member' performance through feedback and coaching. *The Journal of Management Development*, 9(3), 20-27.

Ilgen, D.R., Fisher, C.D., & Taylor, M.S., (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64, 349-371.

Johnson, M., & Helgeson, V.S. (2002). Sex differences in response to evaluative feedback: A field study. *Psychology of Women Quarterly*, 26(3), 242-251.

Kinicki, A.J., Wu, B. Prussia, G.E., & McKee-Ryan, F.M. (2004). A covariance structure analysis of employees' response to performance feedback. *Journal of Applied Psychology*, 89, 1057-1069.

Kluger, A.N., & DeNisi, A. (1996). Effects of feedback interventions on performance: A historical review, a meta-analysis, and preliminary feedback intervention theory. *Psychological Bulletin,* 119, 254-284.

Kramer, M.W. (1995). A longitudinal study of superior^subordinate communication during job transfers. *Human Communication Research*, 22(1), 39-64.

Larson, J.R., Jr. (1984). The performance feedback process: A preliminary model. Or*ganizational Behavior and Human Performance,* 33, 42-76.

Lepper, M.R., Greene, D., & Nisbett, R.E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality & Social Psychology*, 28(1), 129-137.

Lindsley, D.H., Brass, D.J., & Thomas, J.B. (1995). Efficacy-performance spirals: A multilevel perspective. *Academy of Management Review*, 20(3), 645-678.

Locke, E.A., Frederick, E., Lee, C. and Bobko, P. (1984). Effect of self-efficacy, goals and task strategies on task performance. *Journal of Applied Psychology*, 69, 241-251.

London, M. (1988). *Change Agents*: *New roles and innovation strategies for human resource professionals*. San Francisco, CA: Jossey-Bass.

London, M. (2003). Job Feedback: Giving, Seeking, and Using Feedback for Performance Improvement. 2nd edition. Lawrence Erlbaum Associates, Mahwah, NJ.

Maroney, B.P. & Buckley, M.R. (1992). Does research in performance appraisal influence the practice of performance appraisal? Regretfully not. *Personnel Management*, 21(2), 185-196.

Meyer, H.H., Kay, E., & French, J.R. Jr. (1965). Split roles in performance appraisal. *Harvard Business Review*, 43(1), 123-129.

Nadler, D. A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance*, 23, 309-338.

Phillips, J.M., & Gully, S.M. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology*, 82, 792-802.

Podsakoff, M, & Farh, J. (1989). Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior & Human Decision Processes*, 44(1), 45-67.

Russell, J.S., & Goode, D.L. (1988). An analysis of managers' reactions to their own performance appraisal feedback. *Journal of Applied Psychology*, 73(1), 63-67.

Ryan, L.J., & Robey, T.B. (2002). Learning and performance effects of accurate and erroneous knowledge of results on time perception. *Acta Psychologica*, 111, 83-100.

Schleicher, D.J., & Day, D.V. (1998). A cognitive evaluation of frame-of-reference rater training: Content and process issues. *Organizational Behavior and Human Decision Processes*, 73(1), 76-101.

Shrauger, J. S. (1975). Responses to evaluation as a function of initial self-perceptions. *Psychological Bulletin* , 82, 581-596.

Steele, R. & Ovalle, N (1984). Self-appraisal based on supervisory feedback. *Personnel Psychology,* 37, 667-685.

Steelman, L.A. & Rutkowski, K.A. (2004) Moderators of employee reactions to negative feedback. *Journal of Managerial Psychology* , 19(1), 6-18.

Steiner, D.D., Rain, J.S., & Smallery, M.M. (1993). Distributional ratings of performance: Further examination of a new rating format. *Journal of Applied Psychology*, 78, 438.

Stewart-Oaten, A. (1995). Rules and judgments in statistics: three examples. *Ecology,* 76(6), 2001-2009.

Stucke, T.S. (2003). Who's to blame? Narcissism and self-serving attributions following feedback. *European Journal of Personality*, 17(6), 465-478.

Suzuki, K. (1978). Acceptance and rejection of a suggestion. *Japanese Psychological Research*, 20(2), 60-70.

Tang, T., & Sarsfield-Baldwin, L. (1991a). The effects of self-esteem, task label, and performance feedback on goal setting, certainty, and attribution. *Journal of Psychology: Interdisciplinary & Applied*, 125(4), 413-418.

Tang, T., & Sarsfield-Baldwin, L. (1991b). The effects of self-esteem, task label and performance feedback on task linking and intrinsic motivation. *Journal of Social Psychology*, 131(4), 567-572.

Tesser, A., & Rosen, S. (1971). On the reluctance to communicate undesirable messages (the MUM effect): A field study. *Psychological Reports*, 29(2), 651-654/

Van de Vliert, E., Shi, K., Sanders, K., Wang, Y., & Huang, X. (2004). Chinese and Dutch interpretations of supervisory Feedback. *Journal of Cross-Cultural Psychology*, 35(4), 417-435.

Waung, M., & Highhouse, S. (1997). Fear of conflict and empathic buffering: Two explanations for the inflation of performance feedback. *Organizational Behavior & Human Decision Processes*, 71(1), 37-54.

Appendices

Appendix A: Example Norm Table: Positive Sign Condition

Completion Time Norm Data for Model XXG-I17

A student's percentile rank represents the percentage of the norm group that demonstrated a completion time longer than or equal to the individual's completion time. Completion times are to be rounded to the nearest 5 seconds.

Example: Completion of Model XXG-I17 in 7 minutes 58 seconds corresponds to the 40th percentile. This indicates that 40% of the norm group took 8 minutes or longer to complete the model.

Percentile	**20**	**30**	**40**	**50**	**60**	**70**	**80**
Time	**8m45s**	**8m30s**	**8m**	**7m45s**	**7m30s**	**7m15s**	**6m45s**

Note. These norms are highly accurate and were established using a random sample of 378 students, ranging in age from 17-23 years.

© 2002 LEGO Inc. All Rights Reserved. For educational purposes only.

Note. Example for participant with task completion time of 7 minutes and 15 seconds. (Performance time framed as above average, or 70th percentile).

Appendix B. Example Norm Table: Average Sign Condition

Completion Time Norm Data for Model XXG-I17

A student's percentile rank represents the percentage of the norm group that demonstrated a completion time <u>longer than</u> or <u>equal to</u> the individual's completion time. Completion times are to be rounded to the nearest 5 seconds.

Example: Completion of Model XXG-I17 in 7 minutes 31 seconds corresponds to the 40th percentile. This indicates that 40% of the norm group took 7 minutes 30 seconds or <u>longer</u> to complete the model.

Perce ntile	**20**	**30**	**40**	**50**	**60**	**70**	**80**
Time	**8m30s**	**7m45s**	**7m30s**	**7m15s**	**6m45s**	**6m30s**	**6m15s**

Note. These norms are highly accurate and were established using a random sample of 378 students, ranging in age from 17-23 years.

Note. Example for participant with task completion time of 7 minutes and 15 seconds. (completion time framed as average, or 50th percentile).

Appendix C. Example Norm Table: Negative Sign Condition

Completion Time Norm Data for Model XXG-I17

A student's percentile rank represents the percentage of the norm group that demonstrated a completion time longer than or equal to the individual's completion time. Completion times are to be rounded to the nearest 5 seconds.

Example: Completion of Model XXG-I17 in 6 minutes 43 seconds corresponds to the 40th percentile. This indicates that 40% of the norm group took 6 minutes 45 seconds or longer to complete the model.

Percentile	**20**	**30**	**40**	**50**	**60**	**70**	**80**
Time	**8m15s**	**7m15s**	**6m45s**	**6m15s**	**6m**	**5m50s**	**5m15s**

Note. These norms are highly accurate and were established using a random sample of 378 students, ranging in age from 17-23 years.

Note. Example for participant with task completion time of 7 minutes and 15 seconds. (Performance time framed as below average, or 30th percentile).

Appendix D: Feedback Accuracy Manipulation Comments

Condition	Comment
Norm Table is **Accurate**	"This is a LEGO norm table for the model you just completed. It gives you your percentile rank—basically it tells you what percentage of people had a completion time <u>longer</u> than yours—or in other words, what percentage of people you performed <u>better</u> than. The table uses numbers a little differently from grading for a college course, so for example, in a class, somewhere around 70 might be the median score, meaning that ½ the class was higher than 70, and ½ the class was lower than 70. Here, the <u>50</u> is the percentile level for the <u>average</u> performance time. ❖"Your time was **______**" **(show them their time on the timer).** "We tend to round to the nearest 5 seconds, so the table clearly indicates that your performance time falls into the __th percentile of performance, which is **[above/below]** average. So this means that__% of people did better than you and __% of people did worse than you. The table provided by LEGO for this model is based on the performance of hundreds of students and is very accurate. I've seen a lot of people do these tasks, and given the time you took, and the pieces you used, I'd say that I agree with the table."
Norm Table is **Positively** Distorted by Supervisor	"This is a LEGO norm table for the model you just completed. It gives you your percentile rank—basically it tells you what percentage of people had a completion time <u>longer</u> than yours—or in other words, what percentage of people you performed <u>better</u> than. The table uses numbers a little differently from grading for a college course, so for example, in a class, somewhere around 70 might be the median score, meaning that ½ the class was higher than 70, and ½ the class was lower than 70. Here, the <u>50</u> is the percentile level for the <u>average</u> performance time. " ❖ "Your time was **______**" **(show them their time on the timer).** "We tend to round to the nearest 5 seconds, and the table clearly indicates that your performance time falls into the __th percentile of performance, which is [**above/below**] average. So __% of people did better than you, and __% did worse than you. The table provided by LEGO is based on the performance of hundreds of students and is very accurate. **(*Very casually*)** I bet you probably should be rated higher, I kinda liked what you did, I'm willing to say you're probably closer to the __% **(add 15%)** percentile. But then again, the other supervisors in this lab say I'm too lenient and that I give people too high an impression of their performance. Still that is what *I* think."
Norm Table is **Negatively** Distorted by Supervisor	"This is a LEGO norm table for the model you just completed. It gives you your percentile rank—basically it tells you what percentage of people had a completion time <u>longer</u> than yours—or in other words, what percentage of people you performed <u>better</u> than. The table uses numbers a little differently from grading for a college course, so for example, in a class, somewhere around 70 might be the median score, meaning that ½ the class was higher than 70, and ½ the class was lower than 70. Here, the <u>50</u> is the percentile level for the <u>average</u> performance time. " ❖"Your time was **______**" **(show them their time on the timer).** "We tend to round to the nearest 5 seconds, and the table clearly indicates that your performance time falls into the __th percentile of performance, which is **[above/below]** average. So __% of people did better than you, and __% did worse than you. The table provided by LEGO is based on the performance of hundreds of students and is very accurate. ***(Very casually)*** In my opinion, I bet you probably should be rated lower, I watched ya and I'd peg your performance at the __ % **(subtract 15%)** percentile. But then again, the other supervisors in this lab say I'm too strict and that I give people too low of an impression of their performance. Still that really is what *I* think."

Appendix E: Script for Experimental Supervisor Bonding Opportunity

(Participant arrives) 2nd Confederate: "**Hi, are you here for the experiment? What's your name? Good, ok, go ahead and have a seat at the table.** *2nd confederate takes the sign down, puts up "Experiment in session" sign up, and closes the door as participant comes in.* **My name is _______, I'm in charge of this experiment, and this is _________, she is going to be your supervisor for the session today. So for the task that we are trying to validate, there are a couple of models we'll have you build, and in order to get your score on the task, someone has to evaluate your performance, and that is where ___________ comes into play. ________'s been working with us for a while—how long have you been here? I'm not even really sure… well anyway, she has a lot of experience with the task, so you're in good hands. ______, I'm going to take off, is there anything else you need—you have copies of everything you need? Ok well I'm on my way [to a meeting, to class, whatever] so I'll talk to you later."**

Supervisor: "**We'll get started in a minute-- I'm just finishing up a little paperwork here, I have to finish this for [the lab manager, next class, whatever]." (***Shuffle through papers. Wait several seconds, while still "focusing" on your work)* "**So what year are you in undergrad?"** or **"So are you doing anything fun this weekend?"** *Wait for undergrad response, reply to his/her reply.*"

- ❖ Wait a minute before you ask them to help you (because you want to appear as though you are "trying" to get it done yourself as fast as you can)
- ❖ Note what time it is when you start this next section.

Actually, would you mind helping me finish this? It will go more quickly if there are two people working on it, and then we can get started sooner—I want to make sure to get you out in time before the next experiment…. I'm just alphabetizing these articles by author's last name. *(Hand participant stack of papers*)

(*Continue with previous conversation, or continue with questions)* "**So I'm assuming you are taking some psychology courses, are you a major?" "Do you like your classes this semester" "How are your professors this semester?"**

(Once papers are alphabetized, begin experiment) "**Alright, thanks for working on that, let's go ahead and begin the experiment."**

Appendix F: Lego Task Instructions

Lego task instructions. "**Ok, the tasks you will be working on today involve building with Legos. Now even though this is a play toy, we're using it as an indicator of important cognitive, psychomotor and spatial abilities. It also gives an indication of manual dexterity or working with your hands. This building task has been validated many times using younger participants, and we are studying it to see if it is valid for USF students. To make sure it works I'm going to ask you to construct four different models from Legos after a practice session.**

As the supervisor in this study I'll also be able to give you information about how well you have done after the first trial, and after that you are on your own. I will start a stopwatch when you begin each model to determine how long it takes you to complete the task. Keep in mind that you should replicate the model exactly. This means it should match the model piece by piece, including size and color, so whatever lines you see on the model you should replicate -- errors will cost you added penalty time to your completion time.

At the end of each trial I will ask you to answer some questions about your performance, and also at the end of the study. Because I'm interested in getting the most accurate measurement on this building task, please do the best job that you can.

Lastly, there will be a $20 gift certificate awarded to the four participants with the best scores— if you win you'll have your choice of a gift certificate to Chili's, a movie theater, Go-karts, or Maggie Moo's. ***The way we determine who will receive the gift certificates is based on your Lego task-evaluations, which I make based on your overall time for the 4 models.*** **Do you have any questions**?"

Appendix G: Effort Rating

Please circle the number which accurately describes your performance on the task.

On the trial I just completed,

1	2	3	4
I didn't try to improve my performance.	I tried *a little* to improve my performance	I tried *pretty hard* to improve my performance	I tried *as hard as I could* to improve my performance

Note. To be completed after each trial.

Appendix H: Simple Self-Assessment.

Please circle the number that you feel accurately describes your performance on the task.

My performance time on the task was probably better than ___ of other people's performance times.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Below Average					**Average**	**Above Average**				

Note. To be completed after each trial.

Appendix I: Task-Specific Self-Efficacy

Participant # _____

Task-Specific Self-Efficacy

The following scale relates to how you expect to perform on the Lego building task. Please write yes or no in the first column according to whether you think you can accomplish the task listed in each statement. In the second column, for each "yes" please indicate how certain you are of your answer by writing a certainty score from 0 to 100 (0 = completely uncertain, 100 = completely certain).

Item # 1	Column 1	Column 2
I can complete the model in 15 ½ minutes.	_______	_______
I can complete the model in 14 minutes.	_______	_______
I can complete the model in 12 ½ minutes.	_______	_______
I can complete the model in 11 minutes	_______	_______
I can complete the model in 9 ½ minutes.	_______	_______
I can complete the model in 8 minutes.	_______	_______
I can complete the model in 6 ½ minutes.	_______	_______
I can complete the model in 5 minutes.	_______	_______
I can complete the model in 3 ½ minutes.	_______	_______
I can complete the model in 2 minutes.	_______	_______

****

Item # 2	Column 1	Column 2
I can complete the model with 9 or less errors.	_______	_______
I can complete the model with 8 or less errors.	_______	_______
I can complete the model with 7 or less errors.	_______	_______
I can complete the model with 6 or less errors.	_______	_______
I can complete the model with 5 or less errors.	_______	_______
I can complete the model with 4 or less errors.	_______	_______
I can complete the model with 3 or less errors.	_______	_______
I can complete the model with 2 or less errors.	_______	_______
I can complete the model with 1 or less errors.	_______	_______
I can complete the model without any errors.	_______	_______

Note. Completed after each trial. Modeled after Bandura (1977) and Cannon-Bowers (1988).

Appendix J: Participant Standing Sheet

Participant Standing Sheet

We are interested in coaching the supervisor on how to provide feedback. Based on your opinion would you say:

- ❑ My performance feedback ***from the Supervisor*** was accurate and based on the table.
- ❑ My performance feedback ***from the Supervisor*** was too low—my performance was better than she concluded.
- ❑ My performance feedback ***from the Supervisor*** was too high—my performance wasn't as good as she concluded.

Appendix K: Ability Study Questionnaire

Ability Study Questionnaire

(Your responses will not be read by the supervisor, and they do not affect prize eligibility)

1. In my opinion, my performance in trial 1 (Robot) was: (*circle the best answer*)

    1 Below Average (Below 50^{th} percentile)

    2 Average (50^{th} percentile)

    3 Above Average (Above 50^{th} percentile)

---

2. Regarding the feedback you received about your performance after trial one (robot), in your opinion would you say:

- ❑ My performance feedback ***from the Supervisor*** was accurate and based on the table.
- ❑ My performance feedback ***from the Supervisor*** was too low—my performance was better than she concluded.
- ❑ My performance feedback ***from the Supervisor*** was too high—my performance wasn't as good as she concluded.

---

3. After the first trial, when the supervisor gave you feedback on your performance, how much if any did that feedback affect how you performed on trials 2 and 3 (Boat and House)? (*circle the best answer*)

1	2	3	4	5
Not at all	**Slightly**	**A moderate amount**	**A great deal**	**A very great deal**

If the feedback affected you *at all*, did it make you: (*circle the best answer*)

1	2	3	4
Reduce your effort a lot	**Reduce your effort a little**	**Increase your effort a little**	**Increase your effort a lot**

---

4. How often have you played with Legos or other building sets: (*circle the best answer*)

1	2	3	4	5
Never	**Rarely**	**Sometimes**	**Often**	**Very Often**

Appendix K (Continued)

Demographics

5. What is your gender? ______F _____M

6. What is your age? _____

7. What is your ethnicity?

____ African-American ____ White (Non-Hispanic) ____Other

____ Hispanic/Latino/Latina ____ Asian

8. Do you work at a job for pay? ____ Yes ____ No
   If so, do you work part-time or full time? ____ Part-time ____Full-time

Appendix L. Additional Presentation of ANOVA Tables, Means and Plots

Hypothesis 1

ANOVA Results for 1 X 10 Analysis of Hypothesis 1 (Performance)

	Sum of Squares	*df*	Mean Square	*F*	*p*	η_p^2	Observed Power
Intercept	3.25	1	3.25	1784.91	0.00		.92
Condition	0.04	9	0.004	2.336	0.02	0.13	.13
Error	0.27	146	0.002				
Total	3.56	156					

Note. Dependent variable is performance.

Hypothesis 2-4

Estimated Marginal Performance Means[a] and SEs for Individual Conditions and Levels of Independent Variables (ANCOVA Testing of Hypothesis 2)

	Positive Distortion	Accurate	Negative Distortion	Overall Mean for Level of Sign
Positive Sign (70%)	0.166 (0.009)	0.132 (0.010)	0.129 (0.009)	0.142 (0.005)
Average Sign (50%)	0.159 (0.010)	0.143 (0.009)	0.151 (0.010)	0.151 (0.006)
Negative Sign (30%)	0.138 (0.009)	0.149 (0.009)	0.147 (0.009)	0.145 (0.005)
Overall Mean for Level of Distortion	0.155 (0.005)	0.142 (0.005)	0.142 (0.005)	
Overall Mean for Conditions 1-9	0.146 (0.005)			

Note. [a] Means represent transformed completion times.

Explanatory Hypotheses

ANOVA Results for 3 X 3 Analysis of Exploratory Hypothesis (TSSE)

	Sum of Squares	*df*	Mean Square	*F*	*p*	Observed Power
Intercept	977604.75	1	977604.75	6279.89	0.00	1.00
Sign	130.46	2	65.23	0.42	0.66	0.12
Accuracy	92.34	2	46.19	0.30	0.74	0.10
Sign*Accuracy	296.19	4	74.05	0.45	0.75	0.16
Error	22572.49	145	155.67			
Total	1142946.29	155				

Note. Dependent variable is overall TSSE

Estimated Marginal TSSE Means and SEs for Individual Conditions and Levels of Independent Variables

	Positive Distortion	Accurate	Negative Distortion	Overall Mean for Level of Sign
Positive Sign (70%)	87.48 (3.22)	85.55 (3.22)	86.59 (3.12)	86.54 (1.84)
Average Sign (50%)	88.38 (3.34)	82.28 (3.12)	84.77 (3.34)	85.14 (1.89)
Negative Sign (30%)	83.44 (3.03)	85.82 (3.03)	83.39 (3.12)	84.22 (1.77)
Overall Mean for Level of Distortion	86.43 (1.85)	84.55 (1.80)	84.92 (1.84)	
Overall Mean for Conditions 1-9	85.30 (1.83)			

Appendix L (continued)

Follow-up Analyses: TSSE trial 2

Estimated Marginal TSSE Trial 2 Means and SEs for Individual Conditions and Levels of Independent Variables

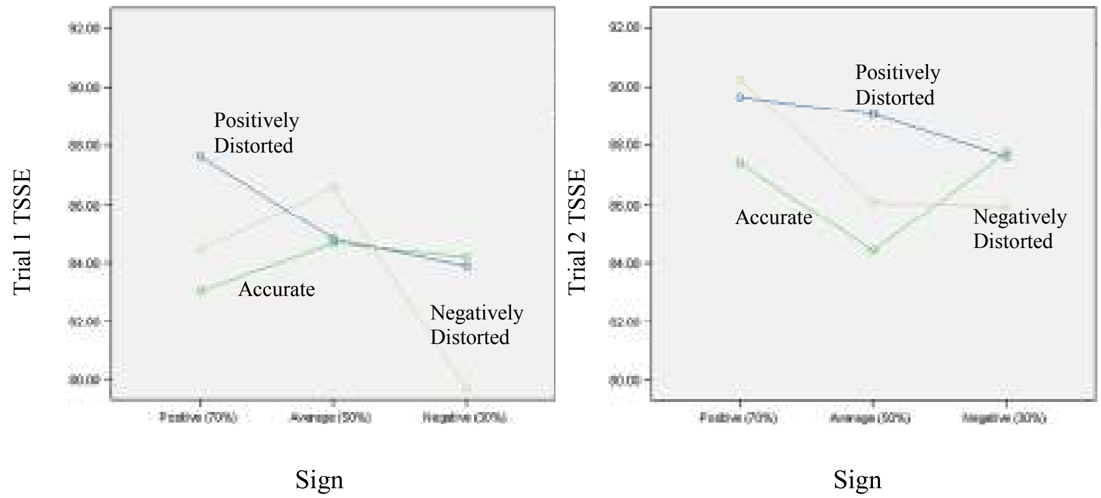
	Positive Distortion	Accurate	Negative Distortion	Overall Mean for Level of Sign
Positive Sign (70%)	89.64 (3.07)	87.40 (3.07)	90.229 (2.97)	89.09 (1.75)
Average Sign (50%)	89.05 (3.18)	84.42 (2.97)	86.07 (3.18)	86.51 (1.79)
Negative Sign (30%)	87.61 (2.88)	87.76 (2.88)	85.93 (2.97)	87.10 (1.68)
Overall Mean for Level of Distortion	88.77 (1.76)	86.53 (1.72)	87.41 (1.76)	
Overall Mean for Conditions 1-9	87.57 (1.74)			

ANOVA Results for 3 X 3 Follow-Up Analysis (TSSE Trial 2)

	Sum of Squares	*df*	Mean Square	*F*	*p*	Observed Power
Intercept	1068456.75	1	1068456.75	7555.97	0.00	1
Sign	166.04	2	83.019	0.587	0.56	0.15
Accuracy	118.86	2	59.43	0.420	0.66	0.12
Sign*Accuracy	156.70	4	39.18	0.277	0.89	0.11
Error	18524.09	131	141.41			
Total	1092241.22	140				

Note. Dependent variable is TSSE trial 2

Appendix L (Continued)



Pre- and post-feedback TSSE reported for trial one and trial two.

Follow-up Analyses: Effort Trial 2

Estimated Marginal Effort Trial 2 Means and SEs for Individual Conditions and Levels of Independent Variables

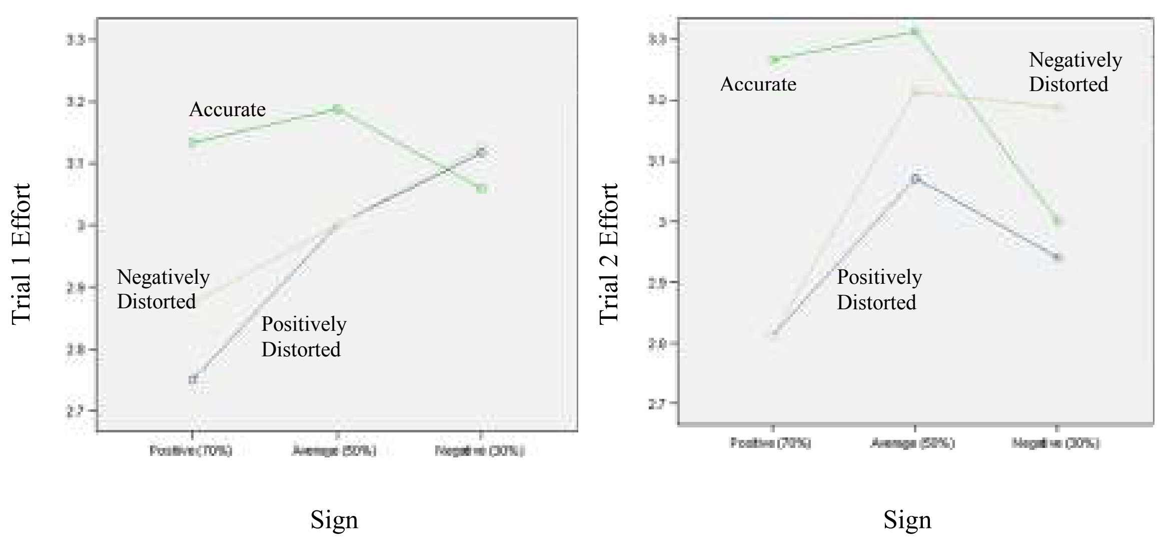
	Positive Distortion	Accurate	Negative Distortion	Overall Mean for Level of Sign
Positive Sign (70%)	2.81 (0.21)	3.27 (0.22)	2.81 (0.21)	2.96 (0.12)
Average Sign (50%)	3.07 (0.23)	3.31 (0.21)	3.21 (0.23)	3.19 (0.13)
Negative Sign (30%)	2.94 (0.21)	3.00 (0.21)	3.19 (0.21)	3.04 (0.12)
Overall Mean for Level of Distortion	2.94 (0.13)	3.19 (0.12)	3.07 (0.13)	
Overall Mean for Conditions 1-9	3.06 (0.13)			

Appendix L (Continued)

ANOVA Results for 3 X 3 Follow-Up Analysis (Effort Trial 2)

	Sum of Squares	*df*	Mean Square	*F*	*p*	Observed Power
Intercept	1321.59	1	1321.59	1818.89	0.00	1.00
Sign	1.29	2	0.65	0.89	0.41	0.20
Accuracy	1.49	2	0.75	1.03	0.36	0.23
Sign*Accuracy	1.68	4	0.42	0.58	0.68	0.19
Error	95.91	132	0.73			
Total	1424.00	141				

Note. Dependent variable is self-report Effort Expended trial 2



Pre- and post-feedback effort reported for trial one and trial two.