RESPONSE TO ASSESSMENT FEEDBACK: THE EFFECTS OF GRADES,

PRAISE, AND SOURCE OF INFORMATION

by

ANASTASIYA A. LIPNEVICH

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Education

written under the direction of

Jeffrey K. Smith

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2007

ABSTRACT OF THE DISSERTATION

Response to Assessment Feedback: The Effects of Grades, Praise, and Source of

Information

By ANASTASIYA A. LIPNEVICH

Dissertation Chairperson:

Jeffrey K. Smith

This study investigated the effects of differential feedback on students' learning progress. The first component of the study consisted of a randomized experiment which involved a large sample of college students working on an authentic learning task (writing an essay). The second component was a series of focus groups with selected participants. The experiment included three conditions: a group receiving no detailed feedback, a group believing their feedback was provided by the course instructor, and a group believing their feedback was computer-generated. Additionally, the three conditions were crossed with two factors of grade (grade/no grade) and praise (praise/no praise) resulting in a 3 x 2 x 2 experimental design. Blocking, based on students' writing ability, was used to assign students to the experimental conditions. After completion of the experiment, eight participants from each condition were recruited for participation in focus group discussions. The purpose of the focus groups was to corroborate, refute, or augment findings from the experiment.

The most pervasive and strongest finding of the study was that descriptive feedback specific to individual work is critical to students' improvement. Although less

detailed forms of feedback, such as grades and general praise messages, led to enhanced

performance, it was to a significantly lesser extent. When a grade from the instructor was

added to descriptive feedback, it depressed students' performance. However, if praise

was also provided, the negative effect of the grade was somewhat ameliorated. Overall,

descriptive feedback was found to be most effective when given alone, unaccompanied

by grades or praise.

Focus group data confirmed the results of the experiment by providing more

information on the nature of differences in students' responses to differential feedback

messages. The significance of the study lies in the application of its findings to classroom

practice. Through better understanding of optimal assessment practices, educators will be

better equipped to support meaningful learning by providing students with feedback that

leads to learning progress.

DEDICATION

No arduous undertaking can reach completion without continuous support, steadfast

encouragement, and difficult sacrifice from all involved. I thank all those who helped me

through this endeavor: those who have steadily believed, constantly reassured,

relentlessly cheered, persistently inspired, patiently tolerated, and unconditionally loved.

This is for you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

Introduction

In the majority of educational programs, a significant portion of the teacher's time

is devoted to the evaluation of students' performance (1988). Ideally, such assessment of

products or behavior should provide students with the necessary information that would

allow them to adjust their learning strategies, get rid of existing misconceptions, correct

mistakes, increase (or sustain) motivation, and, ultimately, move them closer to the

desired goals. However, research has shown that very often the impact of assessment on

students' performance and motivation has the opposite — adverse — effect (Kluger &

DeNisi, 1996; Torrance, 1993). Possible reasons for such a quandary are hidden in

specific attributes of evaluative practices, with the quality and type of feedback being

among the most important ones. Our understanding of differential effects of assessment

in general and feedback in particular on students' learning is crucial to ensure optimal

outcomes of education.

Since research shows that assessment practices may undermine the main goal of

education, that is, to impart learning, then it follows that we should provide teachers with

guidance on how to more appropriately gauge progress and provide meaningful support

to students. The most typical assessment practice is grading. There has been a sustained

controversy in the educational community about the effects of grading on learning

(Guskey & Bailey, 2001; Linn & Miller, 2005). Increasingly, educators claim that grades

should be abolished (Kohn, 1993, 1999; Stiggins, 2004, 2005), however, there are very

few studies to back up this claim. At the same time there are practically no studies

demonstrating a salutary effect of grades as a means to promote learning. Strong claims on both sides of the grading spectrum are based for the most part on speculation.

Likewise, praise is a controversial topic, with some researchers arguing that praise promotes learning by raising positive affect and self-efficacy (Alber & Heward, 2000), while others worry that it leads to depletion of cognitive resources by taking attention away from the task and focusing it on aspects of the self (Baumeister, Hutton, & Cairns, 1990; Kluger & DeNisi, 1996). There is a body of literature providing support for both of these claims (Henderlong & Lepper, 2002). However, this evidence is inconclusive, so the debate continues without a valid consensus from the research community to definitively support either position.

At the same time, computer assisted instruction, use of hypermedia, and sophisticated learning environments have become an ingrained part of modern instructional practices. One of the main functions of many of these complex educational technology systems is to provide students with feedback about their performance. If the effect of teacher-provided feedback seems to be unclear, the impact of computer-provided feedback is even more obscure. There are a few studies showing positive effects of feedback from machines on individuals' performance and affect, but they do not involve authentic learning tasks but rather games or very simple notifications of right-or-wrong responses (Mishra, 2006), and they are mostly conducted in the area of organizational psychology (Earley, 1988; Mishra, Nicholson, & Wojcikiewicz, 2001; Nass, Fogg, & Moon, 1996).

There is ample evidence suggesting that personal dispositions may affect students' reactions to feedback as well as that feedback may affect students' individual

attributes. Studies have examined the role of feedback either influencing students' goal orientation (Latham & Locke, 1991), self-efficacy (Bandura & Locke, 2003), motivation (Kluger & DeNisi, 1996), affect (Ilies & Judge, 2005), or eliciting differential reactions from students depending on those individual characteristics. The findings of these studies, however, have been inconsistent and even contradictory (Kluger & DeNisi, 1996).

In an attempt to fill the existing gap in the research literature, the current study investigated student reactions to differential feedback messages. In a carefully designed experiment using an authentic learning task, the effects of various types of feedback on students' performance and their individual characteristics were investigated. The significance of the study lies not only in advancing our theoretical understanding of how feedback affects individuals, but also in the application of its findings to classroom practices. Through better understanding of optimal assessment practices, educators can be better equipped to support meaningful learning by providing students with feedback appropriate to the educational task and students' individual contexts. This will allow educators to be more effective in helping students develop the skills and knowledge necessary to succeed in future academic, professional, and life settings.

CHAPTER II

Literature Review

From Assessment of Learning to Assessment for Learning

In an argument that changed the conceptualization of assessment, Michael Scriven (1967) suggested differentiation between the summative and formative roles of curriculum evaluation, and it was Bloom, Hastings, and Madaus (1971) who used these terms in their currently accepted meaning. Bloom et al. (1971) defined summative evaluation as tests given at the end of episodes of teaching (courses, units, chapters, etc.) for the purpose of grading, ranking, or certifying students' competence, or for evaluating the effectiveness of curriculum, i.e. to serve the purposes of accountability. This type of assessment practice was juxtaposed with "another type of evaluation which all who are involved — student, teacher, curriculum maker — would welcome because they find it so useful in helping them improve what they wish to do (p. 117)," termed "formative evaluation."

It appears that the difference between formative and summative assessments, suggested by Bloom et al. (1971), resides in the functions they serve, rather than in the nature of these assessments (Black & William, 2003). However, methods used in traditional tests, usually aimed at providing accounts of what has been achieved, might not be very useful in formative assessment, when the teacher is focused on assisting in the learning process. To provide guidance for learning, educators need to employ alternative assessment tools, and should be ready to change their classroom practices (Black & William, 2003).

Torrance (1993), in his analysis of assessment practices, proposed that historically, tests and examinations of all sorts have emerged in response to the social demand to select qualified individuals for educational or work venues, and to predict individuals' future performance based on the evidence from past performance. Such assessment is clearly summative in nature, and has the purpose of measuring achievement and/or ability to be reported to third parties. Current emphasis on accountability, characteristic of the US (Resnick & Resnick, 1991) and European (Krapotkin, Korol, & Elevanova, 2002) educational systems, inadvertently leads to the dominating role of the summative function of evaluation. The main goal of this type of assessment lies in ensuring that learning is happening in classrooms (Linn, 2000). The relationship between assessment and learning, therefore, is thought of as one-directional and very straightforward: You teach something, then you assess it.

However, more recently, reflections on the negative impact of traditional tests have led to an emergence of practical and theoretical work aimed at removing the emphasis from summative assessment and stressing the role of teachers' formative work. For instance, Crooks (1988) notes that "too much emphasis has been placed on the grading function of evaluation, and too little on its role in assisting students to learn" (p. 468). Similarly, Gipps (1994) draws attention to the needed paradigm shift from a testing culture to an assessment culture, and Shinn and Hubbard (1992) contend that there needs to be change from the current assessment paradigm to what they call the problem-solving paradigm, both broadly equivalent to the summative and the formative functions of assessment.

Shinn and Hubbard (1992) argue that the summative functions of assessment are related to consistency of decisions across large groups of students, so that the main requirement is that meanings are shared by different users of assessment results. In contrast, formative functions of assessment highlight desirable consequences for particular individuals, and use the evidence to adapt the teaching work to meet learning needs. Both functions are important, but one does not have to be a savvy educator to realize that if students are to understand and appreciate their strengths and weaknesses and act upon this information to improve learning, carefully tailored assistance must be provided. Formative assessment might be one of the best available forms of such assistance (Black & William, 2003; Fuchs & Fuchs, 1986; Linn & Miller, 2005; Perrenoud, 1998).

Reflective of the above-mentioned view, the term formative assessment has begun to be used interchangeably with the term assessment *for* learning, shifting the frame away from assessment *of* learning (Assessment Reform Group, 1999; Stiggins, 2005). As stated in the American Association for the Advancement of Science report (1998), classroom assessment practices have "the most important direct influence on students' day to day learning" (p. 162). Thus, it seems apparent that for the proponents of formative evaluation, the direction of the relationship between assessment and learning is reversed. It is the information gathered from assessment that is fed back into learning and influences learning, not the other way around.

In his research, Crooks (Crooks, 1988, 2002) has looked into possible ways assessment practices can affect students. In his review of the literature on classroom evaluation, conceptually equivalent to formative assessment, Crooks (1988) presents a

list of ways in which evaluation can influence student outcomes directly and through mediation of numerous personality characteristics. He categorizes the potential effects into three groups of short, medium, and long term effects. Among the possible short term effects of assessment on students he names: (a) focusing attention on important aspects of the subject; (b) encouraging active learning strategies; (c) helping students to monitor their own progress and develop skills of self-evaluation; (d) guiding the choice of further instructional or learning activities to increase mastery, and (e) helping students feel a sense of accomplishment. The medium term effects include (a) influencing students' motivation to study the subject and their capabilities in the subject; (b) communicating and reinforcing the instructor's or curriculum's broad goals for students, including the desired standards of performance; and (c) influencing students' choice of learning strategies and study patterns. Among the long term consequences of assessment are (a) influencing students' ability to retain and apply in varied ways the material learned; (b) influencing students' continuing motivation, both in particular subjects and more generally, and (c) influencing students' self-perceptions (see Crooks, 1989, for the full list of effects, and Gagne, (1977), for a similar list). The proposed categorization is not strict and some influences of assessment obviously extend from short, to medium and long term effects.

An examination of the research on assessment on students makes it clear that the relationship between assessment and learning is close, and that an understanding of the mechanism through which evaluation practices can affect students' achievement is crucial for successful educational practices. In their review of the literature, Black and William (1998) describe the components of formative assessment and the specifics of it.

They propose that the core of the activity of formative assessment is comprised of two types of information and a sequence of two independent but closely related processes. The two kinds of information concern (a) learners' current knowledge set, and (b) the objective knowledge set as prescribed by the instructor, curriculum or students' personal standards. Learners' current knowledge set is reflected by their performance on a given test or task, and can include various forms of knowledge, skills and understanding. The objective knowledge set is represented by the collection of skills and knowledge that the teacher or students themselves have designated as the goal or the main outcome of a particular learning activity. The discrepancy between the two knowledge sets represents a gap which is closed only when the learner has achieved his or her final goal (Black & William, 2003; Ramaprasad, 1983).

The two additional components of formative assessment that Black and William (1998) describe are (a) the perception of learners of a gap between a desired goal and their present state of knowledge, skill or understanding, and (b) the action taken by learners to close the gap in order to achieve the desired outcome. As several researchers note (Black & William, 1998; Torrance, 1993; Tunstall & Gipps, 1996), the key responsibility for generating information about the gap belongs primarily to the teacher whose task it is to discern and interpret the gap and convey this information to the student. However, for familiar or simple tasks with clear requirements and well-understood actions to satisfy them, this responsibility may lie with the student. More complex tasks require higher metacognitive sophistication for students to be able to properly perform self-evaluation, so teachers' involvement in such tasks is imperative (Assessment Reform Group, 1999; Bangert-Drowns, Kulik, & Morgan, 1991).

The action taken by a learner in response to information about the discrepancy depends heavily on the nature of the message, the way in which it was received, the way in which perception of a gap motivates a choice of available courses of action, as well as working contexts in which that action may be carried out (Black & William, 1998). Students' dispositional characteristics, such as their self-efficacy beliefs (Ames, 1992; Craven, Marsh, & Debus, 1991) and goal orientation (Dweck, 1986; Tubbs, Boehne, & Dahl, 1993), as well as temporary affective states (Derryberry, 1991; Ilies & Judge, 2005) are influenced by and, in turn, influence learners' response to the information about the existing discrepancy between the actual and the objective knowledge sets. These and other individual and contextual factors entwined with assessment are addressed in the following sections of the current manuscript.

The research and theoretical work presented above leads to a definition of *formative assessment* as any activity undertaken by teachers, and/or students, which provides information about the gap between students' current knowledge and their objective knowledge, and leads to the subsequent use of this information to decrease this gap, thus furthering learning and promoting academic attainment. This definition covers the important structural elements of formative evaluation — (a) the current knowledge set, (b) the objective knowledge set, (c) gap perception, and (d) gap reduction — and represents a common ground that most researchers in the field of assessment hold (Black & William, 1998; Roos & Hamilton, 2005; Sadler, 1998; Stiggins, 2005; Torrance, 1993).

Consider what appears to be a leitmotif of the previous discussion: In order for assessment to be formative, i.e. to facilitate learning and lead to fulfillment of higher

educational goals, students need to actually *receive* information about their performance and the existing discrepancy between the actual and the desired states of knowledge and process that information. This information is commonly referred to as *feedback* (Ilgen & Davis, 2000; Kluger & DeNisi, 1996).

The two concepts of feedback and formative assessment overlap strongly (Black & William, 1998). The corpus of existing research studies on formative assessment discusses feedback as its inextricable component (Black & William, 1998, 2003; Roos & Hamilton, 2005; Sadler, 1998; Stiggins, 2004, 2005), however, there are many researchers who investigate the concept of feedback without touching upon the notion of formative assessment (Derryberry, 1991; Epsten, Epsten, & Brosvic, 2001). The following sections explore and clarify the construct of feedback in an effort to explicate the nature and characteristics of different types of feedback discussed in the literature. This allows for constraints to be placed on the use of both concepts, providing a more stringent consistency and reaching a level of conceptual clarity imperative for subsequent empirical investigation.

Feedback Research: Historical Perspective

Our understanding of the nature and consequences of instructional feedback has matured considerably over the past decades. Research on feedback dates back almost 100 years, and since then has gone through numerous changes of research and theoretical paradigms. For instance, constructivists discussed the role and nature of feedback, emphasizing the critical function it serves for learning. Piaget defined feedback as the impetus causing disequilibrium, which is followed by accommodation or assimilation of

new information into the individuals' cognitive system. Piaget believed that the acquisition of learning entails an incremental process of change, or development, and that learned abilities are fostered through mediated learning opportunities. Feedback, both internal and external, was perceived as the means through which an advancement of individuals' cognitive functioning occurs (Piaget, 1970/1983.).

Vygotsky (1987) conceptualized feedback as a process of careful scaffolding of student learning and considered it the main tool for helping individuals reach beyond their current cognitive level, that is, to learn and develop (Vygotsky, 1987). He argued that feedback that students receive in the process of assessment "wakens a whole series of functions that are in a stage of maturation lying in the zone of proximal development" (Vygotsky, 1987, p. 212). The constructivist take on the definition and the role of feedback is reminiscent of formative assessment conceptions. In both cases, the ultimate role of external information is to promote learning. Overall, Piaget's and Vygotsky's conceptions of feedback emphasized its educational value, yet did not proffer an answer as to what characteristics a feedback message should have to advance development, neither did they provide a clear framework for carrying out empirical studies to investigate the exact mechanisms through which feedback affects students' learning and development.

Behaviorists had a different stance on feedback. Many of the early studies conducted within the behaviorist perspective viewed feedback as corrective information that strengthened correct responses through reinforcement and weakened incorrect responses through punishment (Kulhavy & Stock, 1989). This mechanistic perspective

stressed the importance of minimizing errors, but provided neither an insight into their correction, nor the means for it.

Feedback following an instructional response was viewed as fitting the sequence of events of the Thorndike's Law of Effect (Thorndike, 1927, 1997), and was construed as the driving force of human learning. The fact that a learner (1) received a task, (2) produced a response, and (3) received feedback indicating whether the answer was correct or not (punishment or reinforcement) provided a superficial parallel to the familiar sequence of the (1) stimulus, (2) response, and (3) reinforcement. However, as Kulhavy and Stock (1989) note, people involved in instructional tasks are not under the powerful stimulus control found in the laboratory, which, along with constantly changing stimuli and responses, bear very little resemblance to the typical operant setting.

Many empirical results were found to be inconsistent with predictions of the Law of Effect (Annett, 1969). For example, Thorndike (1927) postulated that grades, a common example of feedback, can impede learning. He proposed two pernicious properties of such feedback which are its relativity (comparison to others) and indefiniteness (low level of specificity) (Kluger & DeNisi, 1996). These properties, however, could not be explained by the Law of Effect, and made logical and empirical flaws of this paradigm apparent to researchers. Therefore, the reinforcer-punisher view of feedback was deemed too simplistic to account for all the complexities of a learning situation and its consequences on human performance (Anderson, Kulhavy, & Andre, 1971; Annett, 1969).

The cognitive evolution of the late 1960s brought along a shift in the paradigm from feedback as a reinforcer, to feedback as information. A substantial body of evidence

emerged, indicating that feedback operates primarily to correct error responses rather than reinforcing correct answers (Anderson et al., 1971; Kulhavy & Anderson, 1972). Information-processing approaches provided a greater explanatory power in studies of feedback and gradually came to dominate the field. An example of the difference between operant and information-processing approaches lies in the way the two perspectives view feedback about students' erroneous performance.

In the operant approach, presentation of corrective feedback following an incorrect response by definition has no effect on the learner (Kulhavy & Stock, 1989). Thus, errors that students make are ignored, and instructors' attention is directed to students' correct responding only. From the information-processing perspective, on the contrary, errors are of central importance, as this approach describes the exact mechanisms through which external feedback helps to correct mistakes in the products of a learning activity (Kulhavy & Stock, 1989). Overall, information-processing approaches provided tools useful for explaining the impact of feedback on individuals' learning and other feedback-related phenomena that had been theoretically difficult for the behaviorist perspective.

Current theories of feedback have more empirical success in predicting and explaining the impact of feedback on performance and students' individual characteristics. Control theory (Vancouver, 2005), goal setting theory (Locke & Latham, 1990), social cognition theory (Bandura, 1986), and learned helplessness theory (Mikulincer, Glaubman, Ben-Artzi, & Grossman, 1991) include feedback as a central component and have been used to test feedback effects on students' performance and motivation with some success. However, these theories were not general enough to

provide a unified paradigm that would form a conceptual basis for all empirical research on feedback. To deal with this theoretical limitation, Klueger and DeNisi (1996) proposed a unified Feedback Intervention Theory geared at integrating the tenets from existing theoretical paradigms to reach conceptual clarity on many feedback-related constructs, and to explain the observed inconsistencies in the effects of feedback on performance.

A more formal definition of feedback will help set the context for the present consideration of Feedback Intervention Theory, and will help set boundaries throughout the discussion for the phenomenon under investigation.

Operational Definition of Feedback

Definitions of feedback range from simple ones, such as "any information that is provided to the [learner regarding] any action about [their] performance" (Black & William, 1998, p. 47) to more complex ones, such as "information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way" (Ramaprasad, 1983, p. 4). The present manuscript adopts a definition suggested by Klueger and DeNisi (1996) who defined feedback as information regarding some aspects of one's task performance provided by the external agent. The scope of the present investigation is therefore limited to external feedback offered by someone other than the performer him or herself. In an educational context, and in the discussion thereof in the present manuscript, feedback includes written comments made on an essay or a term paper, a grade on a multiple choice test, continuous information of

a learner's correct/incorrect responses during a programmed instruction session, or suggestions made by a teacher while a student works on an in-class assignment.

As previously mentioned, there are currently no sufficiently clear boundaries between the notions of feedback and formative assessment. As a result, the use of these concepts by some researchers lacks consistency (Black & William, 1998). This clarification is proposed to mitigate this conceptual predicament. The relationship between the two concepts can be represented by the following statement: Feedback is used in the realm of formative assessment to shape and improve students' competence.

Any information about products or processes of performance provided to students in a variety of learning situations is, by definition, feedback. Assessment activity, in which this information is intended and is used to facilitate progress in learning, is called formative assessment. Thus, feedback is an essential component of formative assessment and a prime requirement for promotion of students' skills, knowledge and understanding. However, certain types of feedback, such as marking work with a letter grade or providing a total number of test questions answered, have little or no effect on subsequent learners' performance (Crooks, 2002; Tunstall & Gipps, 1996), and therefore do not comply with the main function of formative evaluation — advancement of learning.

Three distinctive features indicating the nature of the relationship between the two concepts can be identified. Firstly, formative assessment and feedback represent two typologically different entities: formative assessment is an activity, whereas feedback is information. Secondly, formative assessment always includes feedback, but there are types of feedback that do not satisfy the main goal of formative assessment, i.e. to facilitate students' academic attainment, and therefore serve different (e.g. purely

summative, or even incidental) functions. Finally, for assessment to be considered

formative, the feedback information has to be used, as feedback per se represents data

that has no necessary consequences (Black & William, 1998; Latham & Locke, 1991).

Feedback Intervention Theory

Feedback Intervention Theory (FIT) (Kluger & DeNisi, 1996) is based on what

the authors refer to as five arguments or assumptions. (1) Human behavior is regulated by

comparison of feedback to goals or standards; (2) goals and standards are organized in a

hierarchical order; (3) only that feedback-goal gap that receives attention can influence

individuals' performance; (4) attention is usually directed to a moderate level of the

hierarchy; and (5) feedback interventions have the capacity to change the locus of

attention within the standard/goal hierarchy. These assumptions are believed to be

interdependent, with every consecutive assumption built upon the previous one.

The first assumption draws upon a common tenet found in both goal-setting

(Latham & Locke, 1991) and control theory (Vancouver, 2005) emphasizing the goal-

directed nature of human behavior. The theories agree that individuals use feedback to

evaluate performance relative to their goals, and that this comparison process produces a

feedback sign — positive or negative. The positive/negative feedback dichotomy is

commonly accepted by proponents of various theoretical orientations (Bandura & Wood,

1989; Carver & Scheier, 1990; Podsakoff & Farh, 1989), with positive feedback being

defined as information communicated to an individual that his or her performance

corresponds to the predetermined standard, and negative feedback construed as

information that individual's performance falls short of his or her desired goal (Vancouver, 2005).

The second FIT assumption proposes that individuals' goals and standards are organized hierarchically. The hierarchies can vary in complexity, but are generally divided into three levels relevant for feedback effects: task-learning processes concerning specific details of a task, task-motivation processes involving the task in general, and meta-task processes involving the self. This proposition draws upon action identification theory which posits that people think about actions in various levels of meaning (Vallacher & Wenger, 1987). For instance, a student may perceive a task of writing a paper as typing words, fulfilling a course requirement, or contributing to personal growth. As individuals become more proficient with a task their attention shifts to higher levels of the hierarchy. This FIT assumption takes us to the next argument which states that individuals may perceive performance-goal discrepancies at all levels of the hierarchy, but only those discrepancies that receive attention are acted upon.

In more technical terms, FIT posits that at the lowest level of the hierarchy there are task-learning processes consisting of the task details, and propositions for improving performance that individuals test by employing feedback. Attention to this level constitutes attention to the details of focal task performance. The next highest level in the hierarchy is comprised of task-motivation processes, which, according to Klueger and DeNisi's fourth assumption, represent individuals' normal locus of attention. At the highest level lie meta-task processes involving the self, affect and general but highly salient individuals' goals. Meta-task processes relate task outcomes to higher-level goals that are important to the individual.

The researchers (Kluger & DeNisi, 1996) propose that a feedback message directing individuals' attention to task-learning or meta-task processes results in a quicker depletion of cognitive resources, whereas messages that focus attention on task-motivation processes lead to the use of universal strategies requiring allocation of little or no additional cognitive resources. Finally, the fifth assumption of the FIT proposes that taking into account all the above information the outcomes of individuals' performance can be influenced by differential feedback messages through means of turning individuals' attention to different levels of the hierarchy. As it was previously indicated, the best outcomes follow the kind of feedback which directs attention towards task-motivation processes (Kluger & DeNisi, 1996; Szalma, 2006).

Although FIT provides a useful framework for understanding feedback and its effects, it seems at times unwieldy and unnecessarily complex, and may not meet all the requirements of a "good scientific theory" (Lerner, 2002; Vancouver, 2005). FIT's abstractions cover a plethora of variables, and the theory appears to coalesce a number of theoretical paradigms, thus imposing a useful structure onto a vast range of conceptions of feedback and organizing our knowledge. However, it has several noteworthy shortcomings. Most importantly, it lacks parsimony. Its attempt to be comprehensive is not balanced enough by critical selections of propositions that this theory is intended to cover. Additionally, the predictive ability of the theory needs to be evaluated through carefully crafted empirical investigations. Klueger's and DeNisi's (1996) attempt to substantiate FIT with empirical findings have lent partial support for the theory. Few studies are available to date that tried to further validate the theory (Szalma, 2006).

Feedback and Performance

Any theoretical perspective that represents learning as a process of interaction and reciprocal influence between students and their environments must accept feedback as a key to improvement of performance and cognitive growth simply because, without feedback, reciprocal influence is by definition impossible (Bangert-Drowns et al., 1991). Exceptions to this rule include evidence of observational, or vicarious, learning (Bandura, 1986) and research on instruction-induced self-questioning (Wong, 1985). Bandura (1986) proposed that students can acquire new behaviors and accumulate knowledge by witnessing persons other than the learners themselves engage in and receive rewards or punishments for their behaviors. Hence, although observational learning is not dependent on special training, its importance is significant for humans who have been taught to profit from the experience of others. Similarly, Wong (1985) showed that some instructional programs designed to elicit students' internal dialogues result in students' self-testing aimed at assessing their own progress. In this case, learners may be capable of providing feedback for themselves without requiring the support of external feedback.

Undeniably, these two approaches can be instrumental in serving an auxiliary role in day-to-day instruction. Yet, the key role of external feedback in providing connections between students' current and desired knowledge states is not to be disputed. This may seem a strong claim to make and to accept, and there are some contingencies to it. For feedback to be effective in facilitating learning progress, it has to satisfy certain requirements, and, as it appears from the literature, far from every kind of external feedback leads to greater learning outcomes (Black & William, 1998; Kluger & DeNisi,

1996). Before exploring the conditions of effective feedback, the multi-faced nature of the phenomenon needs to be considered and a general typology needs to be introduced.

Typologies of Feedback

First, feedback may differ according to intentionality. Intentional feedback is characteristic of instructional settings and is designed to inform students about quality, correctness and general appropriateness of their performance. Unintentional feedback is incidental in nature and results from natural interactions with social and physical environment. The latter often occurs in unstructured peer interactions and unguided simulations (Bangert-Drowns et al., 1991).

Intentional feedback can be further categorized according to the way in which it is provided to students. Direct feedback is delivered from a teacher or a peer to a student in the act of interpersonal communication or in writing. Alternatively, indirect, or mediated feedback, is delivered to learners through a range of artifacts (Leontyev, 1981). Computer-provided feedback is among the most commonly used types of mediated feedback.

Both direct and mediated feedback can be distinguished according to its content on two vectors of *load* and *type of information*. *Load* is represented by the amount of information provided in the feedback message that can range from a letter grade to a detailed narrative account of students' performance (Kulhavy & Stock, 1989). *Type of information* is reflected in the dichotomy of process-related, or descriptive feedback, and outcome-related, or evaluative feedback. Evaluative feedback provides students with information concerning the correctness of responses. It represents a judgment which often

carries a connotation of social comparison (e.g. letter grades, percentile scores, number of solved items, etc.). Descriptive feedback, on the other hand, conveys information about how one performs the task (not necessarily how well) and details possible ways to overcome difficulties with a task and improve performance (Linn & Miller, 2005).

Many researchers proposed alternative typologies of feedback. Bangert-Drowns et al. (1991) suggest that feedback types can be differentiated according to their operation into error correction, presentation of prototypic responses, display of the consequences of responses, and explanation of the appropriateness of responses. Tunstall and Gipps (1996) proposed a more complex categorization of feedback breaking it into two broad categories of feedback as socialization and feedback as assessment. These categories are further broken according to the specific function that a feedback message serves. The lists of functions includes: rewarding/punishing, approving/disapproving, specifying improvements, constructing achievement and constructing the way forward. Construction of a comprehensive typology is beyond the scope of the current manuscript. It is however imperative to understand that feedback is not a unitary phenomenon and indication of the specific kind of feedback should be presented when discussing the effects of feedback on performance and other related variables.

Meta-Analytic Studies of Feedback

Several meta-analyses have been conducted, and extensive reviews of the literature have been compiled that attempted to shed light on the extent of the impact of feedback on students' learning. In their analysis of existing studies, Klueger and DeNisi (1996) present a historical overview of research and show that very often the effect of

feedback on students' learning was judged as unilaterally positive, and evidence that contradicted this assumption was either ignored or deemed to be invalid due to potential study limitations. They contend that flawed methodologies, unwarranted generalizations and empirical inconsistencies of these investigations resulted in a skewed representation of feedback effects on performance and underappreciated variability thereof.

The researchers' meta-analysis (607 effect sizes; 23, 663 observations) demonstrated that feedback improved performance on average (d =.41) but in 1/3 of cases presentation of feedback resulted in decreased performance. This finding cannot be explained by sampling error, feedback sign, or existing theories (Kluger & DeNisi, 1996). The results of moderator analysis showed that (a) feedback effectiveness decreased when individuals received information containing praise or critical judgments which were hypothesized to move students' attention away from the task; (b) correct solution feedback, as opposed to dichotomous judgments of correct/incorrect outcome, led to more effective learning; and (c) effects of feedback on performance on physical tasks was lower than effects of feedback on other tasks. As such, the results of this study provided strong empirical support for the conclusion that feedback is a double-edged sword because it does not always increase performance and under certain conditions is detrimental to performance.

Similarly, the instructional effect of feedback from tests was reviewed by (Bangert-Drowns et al., 1991) using a meta-analysis of 58 experiments obtained from 40 reports. The researchers found that feedback that included any type of elaborated information was consistently more helpful than feedback that informed learners whether

their responses were correct or incorrect. Like Klueger and DeNisi (1996), Bangert-Drowns et al. (1991) revealed variability of feedback effects on performance.

The researchers used further statistical analysis to isolate variables that accounted for the variance in research findings. As a result of this exercise, they found that providing feedback in the form of answers to review questions was effective only when students could not 'look ahead' to the answers before they had attempted the questions themselves, what Bangert-Drowns et al. called 'controlling for pre-search availability.' Controlling for the type of feedback (correct/incorrect versus detailed) and pre-search availability eliminated almost all of the found negative effect sizes, yielding a mean effect size across 30 studies of 0.58.

Two other variables contributed to explaining variance in effect sizes. First, the use of pre-tests lowered effect sizes, possibly by giving learners practice in the material to be covered. Second, the type of instruction determined the effectiveness of feedback, with programmed instruction and simple completion assessment items associated with the smallest effects. Overall, Bangert-Drowns et al. (1991) concluded that the key feature in effective use of feedback is that it must encourage 'mindfulness' in students' responses to the feedback.

The above meta-analyses condense the findings of numerous studies into a rather straightforward conclusion: The effect of feedback on performance is variable and depends on the nature and content of the message provided to a student. This, of course, is a simplified statement, but it guides the further analysis of the literature on feedback, as it reflects the main goal of the current research study — to investigate students' reactions to differential feedback. To rephrase, the pragmatic goal of the present study is to find the

most effective type of feedback to help educators fulfill the main goal of formative

assessment — facilitation of students' learning.

Grading

The most common type of feedback that students receive in a typical classroom is

grades, and more often than not, a letter grade or a numeric score by itself (Marzano,

2000; Oosterhof, 2001). Grades provide a convenient summary of students' performance,

and, with minimal time expense on teachers' behalf, inform all interested parties of

students' achievement. The versatility of the uses of grades is emphasized by many

measurement experts (Airasian, 1994; Marzano, 2000; Nitko & Brookhart, 2007).

Airasian (1994) lists five main functions that grades serve: (1) administrative, dealing

with decisions concerning matriculation, retention and entrance into college; (2)

guidance, by helping counselors provide direction to students; (3) instructional planning,

by informing teachers about students' level of attainment in order to group them for

instruction; (4) feedback, to provide students with information about their progress and

achievement, and (5) motivational, to encourage students to try harder.

If we turn to the previously discussed summative/formative dichotomy of

assessment, it becomes clear that functions of grades 1 through 3 of Airasan's (1994) list

are clearly summative in nature, whereas 4 and 5 are formative. In the former case,

grades are used to inform third parties about students' level of attainment to provide

grounds for making critical educational decisions. In the latter case, grades are provided

to students themselves and are assumed to facilitate students' learning by influencing

their motivation and performance. Although it is hard to disagree with the convenience

and effectiveness of grades when used for summative purposes, the formative function of grades as tools that lead to progress in learning has long been disputed.

One of the main conclusions of the recent review of literature on formative assessment skillfully put together by Black and William (1998) is that descriptive feedback, rather than letter grades or scores, leads to the highest improvements in performance. Moreover, evidence from several studies that investigated the effect of differential feedback on learning suggests that using grades to improve learning is a dubious practice.

To investigate the effect of grades on students' performance and motivation, Butler and Nisan (1986) conducted a study involving 261 sixth-grade students. Students were presented with two tasks, one quantitative and one qualitative, on three sessions over 2 days, and were randomly assigned to three feedback conditions of no feedback, grades only and individualized comments which consisted of one phrase relating to some aspect of the task that the child had performed well, and one phrase relating to an aspect which could have been improved. Feedback was manipulated after the first and the second experimental sessions, and measures of students' performance consisted of the third session performance scores and self-reported motivation.

The analysis revealed the following performance patterns: The individual comments group scored high on both tasks, whereas the no-feedback group scored low, showing a significant decline in scores from pretest. Students who received grades only scored high on the quantitative task and scored lower than at pretest on the divergent thinking measures. In terms of feedback effect on motivation, Butler's and Nisan's (1986) hypothesis that intrinsic motivation would be maintained after receipt of task-

related evaluation and undermined after non-receipt of feedback or receipt of normative grades was supported. Additionally, those subjects who received written comments expressed greater interest in the tasks than did those in the other two conditions, especially for the question requiring greatest commitment — the number of extra tasks chosen (Butler & Nisan, 1986).

In their discussion of the results of the study, the authors stipulate that grades tend to emphasize quantitative aspects of learning, depress creativity, foster fear of failure, and weaken students' interest. Quite opposite to this pattern, no negative consequences follow from the use of task-specific individualized comments. The authors state that varied effects of these types of assessment stem from the salient control aspect of grades, as well as from differences in the information both modes of evaluation provide: Normative grades inform students' about proficiency relative to others, whereas individualized comments create clear standards for self-evaluation specific for the task (Butler & Nisan, 1986).

In a study by Elawar and Corno (1985), the effect of teachers' written feedback provided to students' homework was investigated with over 500 Venezuelan students involving 18 mathematics teachers in three schools. The researchers trained the teachers to give written feedback which concentrated on errors and on poor strategy, followed by suggestions about how to improve performance on future tasks. The descriptive personalized feedback was provided to students in the experimental condition. For the control group, teachers were instructed to follow the normal practice of assigning grades without comments. Additionally, a third group of the trained teachers marked half of their classes with full feedback and the other half with letter grades only. All students were

given a pre-test and one of three parallel forms of post-test. Analysis of the results demonstrated a large effect associated with the feedback treatment, which accounted for 24% of the variance in final achievement. Students who received comments performed significantly better then those who received grades. The authors explain the findings through the lens of cognitive theory and research, which emphasizes the importance of deep processing when acquiring complex information. Apparently, comments provided by teachers assisted in turning students' attention to relevant, specific information, stimulated mental elaboration and as a result boosted their performance. Grades, perceived as reinforcers and punishers, which are believed to be controlling and lacking specificity, led to inhibition of students' cognitive processes and slower progress of learning.

Since the prevalence of grades in the current education system is apparent, finding ways of using them effectively would make an immense contribution to assessment in education. Led by this goal and in an attempt to further our knowledge on the effect of grades on students' performance and motivation, Butler (1988) conducted a follow-up experiment that involved 48 11-year-old Israeli students, with half being of high and half being of low academic ability. The students were encouraged to work on two types of tasks individually under supervision, with one task testing convergent thinking, and the other one divergent. The procedures were similar to those of the earlier study; feedback manipulations, however, were different.

To capture the effect of feedback, and to provide a more complex picture of possible consequences of grading, Butler (1988) derived the following experimental conditions. The first group received individually composed comments indicating the

match, or not, of their work to the criteria specified beforehand in the study instructions. Students in the second group were given only grades, and the third group was presented with both grades and comments. For every student, scores reflecting their performance in each of the three sessions were calculated and served as outcome measures.

Consistent with the results of the Butler and Nisan (1986) study, the results of the inquiry showed that the group that received comments specifically tailored to students' performance showed a significant increase in scores (by almost 30%) for both tasks between the first and second sessions, and remained at this high level for the third session. The group that received grades only showed a significant decline in scores on both tasks, and, so did the group that received both grades and comments. Analysis of students' reports of interest in performing the task demonstrated a similar pattern, with interest being undermined for both graded conditions. Interestingly, high achievers in all three feedback regimes sustained a high level of interest, whereas low achievers in the graded groups reported a dramatic decline thereof (Butler, 1988).

The researcher discusses these results in terms of cognitive evaluation theory, and posits that even if feedback comments are helpful for students' work, their effect can be undermined by the negative motivational effects of the normative feedback, that is by giving grades and scores. Butler (1988) proposes that a general preoccupation with grade attainment often lowers the quality of task performance, especially on more complex tasks.

Several studies investigating the impact of grades on students' learning present evidence which is in agreement with Butler's (1988, Butler & Nisan, 1986) findings. For example, in an experiment conducted by Grolnick and Ryan (1987), those students who

were told they would be graded on how well they learned a social studies lesson had more trouble understanding the main point of the text than did students who were told that no grades would be involved. Even on a measure of rote recall, the graded group remembered fewer facts a week later. Another study presented the evidence that students who tended to think about the material they study in terms of what they would need to know for a grade were less knowledgeable than their counterparts, even after taking other variables into account (Anderman & Johnston, 1998).

In addition to the motivational theory explanations, the negative impact of grades on students' performance can be explained by Feedback Intervention Theory (Kluger & DeNisi, 1996) described in the preceding section of the present manuscript. This theory suggests that the optimal feedback should direct individuals' attention to the details of a specific task and to learning methods that would help achieve desired results. Based on this logic, letter grades and numerical scores would tend to channel students' attention to the self, and away from the task, thus leading to negative effects on performance (Siero & Van Oudenhoven, 1995; Szalma, 2006; Szalma, Hancock, Warm, Dember, & Parsons, in press).

Overall, many researchers are quite persuasive in their criticism of letter grades or any form of quantitative evaluation of students' performance (Linn & Miller, 2005; Roos & Hamilton, 2005; Smith, Smith, & DeLisi, 2001; Stiggins, 2004). In his summary of the pernicious effects of grades on students' learning, Kohn (1993) claims that grades tend to reduce students' interest in the learning itself, and backs up this claim by citing numerous studies from many domains showing that the more people are rewarded for doing something, the more they tend to lose interest in whatever they had to do to get the

reward. Additionally, he proposes that grades tend to reduce students' preference for challenging tasks, as students' would naturally try to obtain high grades with minimal time, effort and intellectual expenses. The reduction of the quality of students' thinking is listed as yet another detrimental consequence of assigning grades (Kohn, 1993, 1999; Linn & Miller, 2005).

This strictly negative position on the detrimental role of grades on students' performance may be prevalent, but it is not the only one in the field of assessment. In an attempt to refute a commonly voiced urge to abolish grades, Marzano (2000) states that the most important purpose for grades is to provide feedback to students and, if referencing for grading is content-specific, letter grades and numerical scores will lead to an increase in students' performance. He postulates that if students have a clear understanding of the requirements of the task and if grading is based on students' achievement and effort only, students can increase their level of knowledge and understanding based on grades alone.

Guskey and Bailey (2001) take a similar stance on the issue of grades. They suggest that if grading is done properly, an increase in students' academic attainment will follow. To back up their argument, the authors describe a study conducted by Page (1958). This work is considered to be classical and is often cited by scholars who do not belong to the radical grade-refuting camp. In his study, Page (1958) had 74 secondary school teachers administer a test to students in their classes and provide feedback according to the condition to which students were randomly assigned. The first group received a numerical score along with the corresponding grade. The second group received a score along with a standard comment ranging from "Excellent! Keep it up!"

(accompanying an A) to "Let's raise this grade!" (following an F). For the third group, teachers were asked to provide extensive individualized comments along with the grade. The analysis showed that students who received individualized comments in addition to a numerical score and a grade outperformed the other two groups. Additionally, students who received a grade followed by a standard comments performed significantly better than students in the grade only group. Based on these results Page (1958) concluded that grades can be effective for promoting students' learning when accompanied by a comment.

Guskey and Bailey (2001) discuss the findings of Page's study in order to convince the reader that grading can be used quite effectively to enhance students' academic achievement. They go further and propose that the beneficial effects of grades "can be gained with relatively little effort on the part of teachers. Stamps, or stickers with standard comments such as these [described in the study] could be easily produced for teachers' use" (Guskey & Bailey, 2001, p. 29). Obviously, this claim is likely to be deemed too simplistic by many researchers in the field, especially if we take into account the fact that Guskey and Bailey (2001) drew their conclusion from a sole study conducted half a century ago.

Overall, the review of the studies on grading reveals a lack of current inquiries into the effects of grades on students' learning. Most of the existing studies are dated and have a variety of methodological flaws. To list a few, Page's (1958) study did not include either a comment only group or a no-feedback control group, and comments provided to students were not meaningful in terms of analyzing their performance or providing guidance for improvement. Should these shortcomings have been accounted for, it is

quite possible that the outcome of the study would have been different. Butler's studies (Butler, 1988; Butler & Nisan, 1986) lack ecological validity because the experimental tasks were not part of or related to normal curriculum work and were not carried out by the students' regular teachers. Elawar's and Corno's (1985) study was carried out in Venezuela, and written comments were provided in Spanish. Possible issues with inherent cultural differences concerning school values, teacher-student relationships, as well as specifics of the Spanish language should make us cautious when using the results of this study to substantiate strong claims in the ongoing debate about the effects of grading. Most importantly, all of the cited studies were conducted with school children of younger age, older students do not appear to have been a target of such inquiries.

The combination of these listed deficiencies strongly suggested that a new study was in order that would adequately address the existing need to clarify and expand our knowledge on the effects of grading on students' learning. In a randomized experiment that involved a large sample of college students, followed by a series of focus groups with selected participants, this study attempted to account for the shortcomings of the previous studies and investigated the effect of grades on students' progress of learning.

Praise

The previous section briefly touched upon a type of feedback, commonly referred to as praise. Praise has been defined as "favorable interpersonal feedback" (Baumeister et al., 1990, p. 131), or "positive evaluations made by a person of another's products, performances or attributes " (Kanouse, Gumpert, & Canavan-Gumpert, 1981, p. 98). This type of feedback is probably the second most common kind (with the first being grades)

that students' receive from their teachers, and it runs the gamut from simple "You did a great job!" statements to much more elaborate and personalized positive references to students' performance. Generally, praise is believed to have beneficial effects on students' self-esteem, motivation, and performance. As a result, teachers are encouraged to use praise as a reinforcer of a desired behavior (Dev, 1997). However, quite similarly to the research on grading, the conclusions concerning the impact of praise on students' performance lack consistency.

There are two opposing views on the effect of praise on students' learning. One camp of researchers and educators claims that normally, a feedback message containing praise enhances intrinsic motivation and leads to improvement in individuals' performance (Cameron & Pierce, 1994; Dev, 1997; Pintrich & Schunk, 2002). Shanab and colleagues (1981) investigated the influence of praise on motivation, operationalized through interest and persistence. They found that praise during a puzzle-solving task led undergraduates to spend more time on the task and to rate their interest as higher than participants in a control condition who received neutral feedback. Similarly, meta-analytic studies examining the effects of praise on motivation have shown that positive statements have a tendency to increase intrinsic motivation across a variety of dependent measures (Cameron & Pierce, 1994; Deci, Koestner, & Ryan, 1999). This effect, however, is not always strong, varies for different age groups and was often derived in the course of methodologically flawed studies (Henderlong & Lepper, 2002; Lepper, Henderlong, & Gingras, 1999).

The researchers who emphasize the positive role of praise for students' learning refer to a number of theoretical mechanisms to explain there results. One commonly

discussed variable which is believed to mediate the effect of praise is self-efficacy, defined as the belief that one has the capabilities to execute the courses of actions required to achieve desired outcomes (Bandura, 1997; Bandura & Locke, 2003). Drawing upon a long line of research, Bandura (1986, 1997) proposed that individuals' self-efficacy is strongest when it arises from their own achievement, but persuasion can be effective in convincing individuals that they have the ability to succeed. So, in this circular process praise can be used to make students believe that they can succeed, which should, in turn, enhance self-perceptions of efficacy and lead to greater academic attainment.

Feedback containing praise may also be effective because it elicits a positive affective reaction, which has been often linked to increased motivation and higher goals (Delin & Baumeister, 1994; Ilies & Judge, 2005). This mediating role of affect in influencing individuals' behavior can be explained with Gray's behavioral motivation theory (Gray, 1990). Gray's theory suggests that there are two distinct systems that regulate motivation. The first one is the behavioral activation system (BAS), which is believed to regulate appetitive motivation and is activated by stimuli signaling reward (or relief from punishment). The second one is the behavioral inhibition system (BIS), which regulates aversive motivation and is activated by stimuli signaling punishment (Gray, 1990). BAS is believed to regulate the experience of positive emotions and moods, whereas the BIS controls regulation of negative emotions and moods.

Gray (1990) proposed that stimuli from the environment influence people's affective states, and that resulting affective states will reinforce behavioral motivation. For example, because positive affect which often follows praise has an energetic arousal

component, it should increase individuals' optimism concerning performance, and thus cause increase in effort and persistence. Drawing upon Gray's theory, Ilies and Judge (2005) propose that favorable feedback cues would directly lead to positive affect, which is associated with BAS activation, so individuals will engage in approach behaviors and set higher goals as a result of it. Ilies and Judge (2005) conducted a series of experiments using six independent samples totaling more than 900 participants, with three types of tasks, and with three types of performance feedback that demonstrated that basic affective reactions to feedback are important mechanisms that explain the relationship between feedback and future goals.

Another explanation of the positive effect of praise on behavior is proposed by Henderlong and Lepper (2002). They posit that children may continue to exhibit praised behavior to sustain the attention and approval of the evaluator because of the positive interpersonal dynamic that typically characterizes occurrences of praise. They note, however, that motivational benefits may be purely extrinsic and quite transient, dissipating as soon as the evaluator is no longer present (Henderlong & Lepper, 2002).

Finally, the mechanism through which praise is believed to influence learning is often borrowed from the behaviorist literature. Behavior modification programs are developed that emphasize the systematic and contingent use of praise over time for the purpose of reducing classroom behavior problems and encouraging students to learn. Studies in the behavioral tradition have shown that praise can be indeed a successful technique for influencing a broad range of students' classroom behaviors (Alber & Heward, 1997, 2000; O'Leary & O'Leary, 1977). However, studies that employed behavior modification techniques seem to have a common problem: Despite the fact that

they demonstrate success of positively stated feedback, praise is almost never isolated as a single variable. As Henderlong and Lepper (2002) note, the effects of praise in such studies is often confounded with numerous contextual variables, and therefore should be judged with care.

Evidence of a direct or mediated positive influence of praise on motivation and performance is not without flaws, but is abundant. It is apparent that there are many plausible mechanisms that may potentially account for such effects, but these mechanisms should be subjected to more careful examination. There are also examples of the negative impact of praise on students' learning. A good starting point might be Baumeister's et al. (1990) study, which presents evidence that praise can both impede and facilitate individuals' performance.

Baumeister et al. (1990) conducted four consecutive experiments that involved a total of 172 undergraduate students working on two tasks: one requiring pure effort for successful completion thereof (sorting cards), and the other one requiring skilled performance (playing a video game). Praise was presented to the participants in the experimental condition, and change in performance from the baseline trials to the trial following the praise was used as a dependent variable. In the control condition, the experimenter made a neutral comment irrelevant to the task. The analyses showed that positively-framed feedback improved students' performance on a pure effort task, but consistently led to impairment in skilled performance. Additionally, the researchers found that both task-relevant and task-irrelevant praise resulted in performance decrements. When discussing these results, the authors quite humorously note that "an effective way

to disrupt skilled performance is to compliment the performer immediately beforehand"
(Baumeister et al., 1990, p. 145).

On a more serious note, the researchers proposed three possible mechanisms by
which praise could impede successful task completion. The first is that praise may cause
participants to feel that they no longer need to exert as much effort on further trials. This
explanation was repudiated by the finding that praise improved performance on the
mechanical card-sorting task. The second explanation was that task-relevant praise
carries a demand for good subsequent performance thus creating pressure. This
explanation received partial support, but had difficulty accounting for the damaging
effect of task-irrelevant praise. Finally, the third proposed mechanism is that praise
makes individuals self-conscious and leads to disruption of skilled performance. The fact
that task-irrelevant praise negatively affected task outcomes provided further support for
the latter explanation — it is hard to explain otherwise why comments about hair or outfit
would lead to a decrease in performance. Apparently, attention to the self resulting from
praise robs cognitive resources that would otherwise be committed to the task. Only if a
task is automated, and fewer resources are needed for its completion, would praise have a
neutral or positive effect on performance. Therefore, the assumption that praise focuses
attention on self, and not the task, seems to be the most plausible and parsimonious in
explaining the negative effect of praise on performance. It is also in accord with the
tenets of FIT proposed by Klueger and DeNisi (1996).

Baumeister's et al. (1990) study provides valuable data which helps explain the
negative impact of praise on students' performance, however, it did not include a no-
praise control group, making it difficult to rule out alternative explanations of the

findings. Boulet, Simard, and Demelo (1990) and Butler (1987) accounted for this shortcoming and included a no-praise condition into their experiments. To examine the impact of praise on students' motivation and performance, Boulet et al. (1990) conducted an investigation with a group of 80 Canadian high school students. The participants were randomly assigned to one of the two experimental groups or the control group for a course on the writing of the major scales in music. Students in the first experimental group were given feedback on their pre-test in the form of written praise, a list of weaknesses and a work plan for further instruction, whereas the second experimental group received oral feedback, were told about their errors and given the opportunity to correct them. The analysis of students' performance revealed significant gains by the second experimental group as compared to the first experimental group or the control group. The authors' interpretation of the findings suggests that the oral delivery of feedback is more effective than written delivery of feedback. However, drawing from the previous discussion it seems more probable that the positively-laden message which was presented in the written feedback directed students' attention towards their selves, rather than towards the specifics of the tasks (Baumeister et al., 1990; Black & William, 1998; Kluger & DeNisi, 1996).

Additional evidence of the negative effect of directing students towards the self rather than the task comes from a study carried out by Butler (1987). In it, she examined the effects of comments, grades, praise, and no feedback on the performance outcome of 200 Israeli students. Despite the fact that the four groups were matched on pre-test scores, the students who received comments scored one standard deviation higher than the other groups on the post-test. No significant differences were detected among the other three

groups. Additionally, students given grades and praise scored far higher than those receiving comments or no feedback at all on measures of ego-involvement, whereas those given comments scored higher than the other three groups on measures of task-involvement. Furthermore, students in the praise condition had the highest perceptions of success, even though they had been significantly less successful than the comments-receiving group.

In sum, there is ample evidence providing support for claims at both ends of the praise spectrum. However, this evidence is inconclusive, and new studies that carefully examine the effect of positively-framed feedback would make a valuable contribution to the field. The present study attempted to accomplish this goal. In order to enhance ecological validity, the study involved a task which was a part of a regular undergraduate course requirement. This accounted for the major shortcoming of the above-cited studies, as most of them employed tasks which were not part of routine educational practices (Baumeister et al., 1990; Butler, 1987; Ilies & Judge, 2005).

Additionally, in many cases, drawing conclusions about the effects of praise was complicated by the absence of a control group and numerous confounding variables present in the design of the studies (Baumeister et al., 1990; Boulet et al., 1990; O'Leary & O'Leary, 1977). The present research included a control group receiving no detailed feedback, and two experimental groups which were presented with individualized feedback coming from either computer or the course instructor. The three conditions were crossed with the grade and praise factors. Thus, for half of the students in each condition laudatory statements on the performance were included. Presence of these

conditions allowed for teasing out and evaluating possible effects of praise on students' performance.

Source of Feedback

The typology of feedback provided elsewhere included a dichotomy of direct versus mediated feedback. Computer assisted instruction, use of hypermedia, and sophisticated learning environments have become an ingrained part of modern instructional practices. One of the main functions of many of these complex educational technology systems is to provide students with feedback about their performance. If the effect of teacher-provided feedback seems to be unclear, the impact of computer-provided feedback is even more obscure.

Researchers investigating the nature of human-computer interaction can be divided into two camps. The first group believes that people tend to view computers as neutral tools that bypass issues of attitude, affect and stereotypes characteristic of human interactions. These scholars posit that computer-provided feedback will elicit individuals' reaction different from the one following human-provided feedback (Lajoie & Derry, 1993; Lepper, Woolverton, Mumme, & Gurtner, 1993). Furthermore, researchers in this paradigm state that users and learners will tend to be skeptical towards computer-provided "personal" comments, and will find computer responses such as praise, criticism and helping behavior as being implausible and unacceptable (Lepper et al., 1993).

The other group takes a different stance on the matter. These researchers describe themselves as functioning within the Computers as Social Actors paradigm, and

demonstrate that people may be unconsciously perceiving computers and other media as being "intentional social agents" (Nass, Moon, & Carney, 1999). Some studies show that people often attribute human characteristics to computers: People are polite to machines (Nass et al., 1999), perceive machines as competent teammates (Nass et al., 1996), ascribe gender and personalities to machines (Nass, Moon, & Green, 1997), and get angry and punish them (Ferdig & Mishra, 2004). Responding socially to a computer is also quite common, and typical for people of all ages and levels of expertise (Mishra, 2006). People are found to talk to computers even though they explicitly deny believing that computers have feelings or intentionality (Reeves & Nass, 1996). Therefore, the supporters of the CASA framework would propose that human and computer-provided feedback would have the same or very similar effect on individuals.

Studies that examined the impact of computer versus human-provided feedback are few and far between, and are mostly conducted in the stream of organizational psychology research. The results obtained in the course of those studies are quite intriguing for the realm of education and may have important implications. For instance, a study conducted by Earley (1988) inquired into a contrast between computerized feedback and feedback provided by the supervisor in a subscription-processing job. The results showed that computerized feedback was more trusted, led to stronger feelings of self-efficacy, to more strategy development, and to better performance compared with identical feedback coming from a supervisor. These findings seem to support the argument of those researchers who believe that computers are perceived by individuals as neutral tools, and consequently, unbiased sources of information. Because machines do not elicit any affective responses from individuals, cognitive resources get directed

towards tasks resulting in an increase in performance. The results can also be explained with Feedback Intervention Theory (Kluger & DeNisi, 1996). Feedback provided by the supervisor could have directed participants' attention to meta-task processes, such as evaluating the intentions of the supervisor and their implications for goals of the self, whereas the computerized feedback directed attention to the task and to the task details.

A more recent study was conducted by Mishra (2006) who investigated the effects of feedback provided by computer. The design of Mishra's study mimicked the procedures of a study carried out by Meyer, Mittag, and Engler (1986) which examined the effect of experimenter provided feedback on participants' self-evaluations of performance and affect. In the Meyer et al. (1986) experiment, students were asked to take a test of logical ability, after which the experimenters pretended to score tests for half of the participants, and the other half was asked to proceed directly with two tasks: one easy and one difficult. This manipulation was supposed to induce the students to believe that the experimenters knew their ability in one case, and did not know it in the other. After completion of each task students were provided with either laudatory or critical comments. In sum, the results of the study showed that informed praise was counterproductive when offered for success on an easy task, and criticism increased motivation when presented after failure on a difficult task. When the evaluators had no knowledge of the participant's ability (i.e. in the non-test-scored condition), the feedback had no significant effect on any of the performance evaluations or affective reactions.

Analysis of the results obtained in Mishra's (2006) study showed that computer-provided feedback also made a significant difference to the participants' motivation and affect; however, the pattern of responses was different from the one described in Meyer

et al. (1986). The participants who received positive feedback for success on both difficult and easy tasks and neutral feedback on failure on a difficult task, had more positive affect in regards to their performance, rated their performance as being better than those who received neutral feedback for success on the easy task and blame feedback for failure on the difficult task, and perceived the evaluation of the computer as being more fair than those who received negative feedback. No differences were found between the participants who believed that their initial tests of logical ability were scored and those whose tests were not scored. Overall, praise had a uniform positive impact on participants' motivation and affect — an effect different from the results obtained in Meyer and colleagues' study.

In the discussion of the results Mishra (2006) states that if the "computers as neutral tools" paradigm is true no difference between the groups based on the experimental manipulations should have been found. The fact that participants did respond to feedback from computers and that it did make a difference to their self-perception and motivation provides support for the alternative view — the Computer as Social Actors paradigm. The support, however, is only partial because the obtained results were different from the human-human study. To explain this discrepancy Mishra (2006) proposed that people accept feedback from the computer at face value without reading into the context of the feedback.

Mishra's (2006) study gives us initial answers to questions concerning individuals' reaction to computer-provided feedback. It shows that students do form affective reactions towards feedback provided by the machine, but the nature of the differences between their reactions to computer-provided feedback and their reactions

toward human-provided feedback remains unclear. The fact that the researcher compared the results of two studies conducted in different times and places and involving different populations, makes the inferences about these comparisons somewhat questionable. Additionally, the impact of computer-provided feedback on individuals' performance and motivation is still not clearly understood.

Earley's (1988) study provided initial information concerning the differential influence of human and computer-presented feedback messages on performance, self-efficacy and strategy development. The experimental task and population of adult professionals, however, does not allow for the generalization of these results to the domain of education. Therefore, a study conducted in an educational setting involving students working on a relevant task is in order.

To fill the gap in the existing literature on computer versus human-provided feedback the present study examined the effect of detailed constructive feedback provided by the course instructor and by a computer program on students' learning. In two experimental conditions students received the same type of descriptive, individually-tailored feedback, and their perception of the source of feedback was manipulated: One group was informed that the comments were generated by the computer program, and the other group was induced to believe that the comments came from the course instructor. Unlike Mishra (2006) who inquired into affective reactions to computer-provided comments, this study attempted to uncover motivational and performance-related effects as well. The results obtained in this study have important implications for educational technology and may potentially influence views of designers of educational software.

Additionally, it provided more information for the ongoing debate about the nature of human-computer interactions.

Individual Characteristics

The research literature reviewed so far suggests that individual characteristics, such as self-efficacy, type of goal orientation, ability level, and level of self-regulation may contribute to differential responses to feedback messages.

Goal Orientation

Goal orientation has been consistently found to have an effect on students' learning, and it has been suggested that the type of goals that individuals have may lead to variation in their responses to feedback (Black & William, 1998; Ilgen & Davis, 2000). Goal orientation is the central concept of achievement goal theory — one of the most applicable paradigms used to understand students' academic motivation. This theory states that individuals engage in academic tasks to fulfill different goals that can be broadly broken into two categories of performance and mastery (Ames, 1992; Mattern, 2005; Schunk, 1990).

Performance goals drive individuals to seek and maintain a positive image of their ability (Pintrich & Schunk, 2002). In early versions of the achievement goals theory, performance goals as a whole were seen as being maladaptive for learning (Dweck, 1986). However, recent studies have differentiated the construct of performance goals into approach and avoidance. Individuals with performance approach goal orientation are characterized as being motivated to outperform others, to demonstrate their superiority,

and to attain favorable judgments of competence from teachers, parents, and peers, whereas individuals with performance avoidance goal orientation strive to avoid failure or to appear incompetent in comparison to others (Elliot, 2005; Pintrich & Schunk, 2002). Research evidence from these studies suggests that performance approach goals are related to more positive outcomes, such as use of cognitive strategies (Pintrich & Zusho, 2002; Pintrich, 2000), and improved performance (Church, Elliot, & Gable, 2001; Harackiewicz, Barron, Elliot, Carter, & Lehto, 1997), whereas performance-avoidance goals are related to superficial learning strategies, lower levels of achievement, low interest and decreased intrinsic motivation (Archer, 1994; Luu, Tucker, Derryberry, Reed, & Poulsen, 2003). Mastery goals, on the other hand, direct individuals to seek opportunities to reach new levels of competence and master complex tasks (Dweck, 2000). These students tend to seek more challenges, hold positive attitudes towards school, and report a higher level of self-efficacy than those individuals who pursue performance goals (Ames, 1992; Midgley, Kaplan, & Middleton, 2001; Schunk, 1990; Wolters, 2004).

Educational researchers commonly agree that mastery orientation needs to be encouraged for students to be active and effective learners. Yet, there are situations in which a performance orientation may actually lead to greater improvement in performance and more efficient use of sophisticated cognitive strategies (Wolters, Yu, & Pintrich, 1996). For instance, when students are presented with learning tasks that do not capture their interest or fail to challenge them, focusing on outperforming others may make the learning task less boring and, as a result, could lead the student to employ self-regulatory skills and result in higher attainment (Boekaerts & Niemivirta, 2000).

Several studies have demonstrated that feedback, especially feedback indicating poor performance, is interpreted very differently depending on the goal orientation (Ames, 1992). When individuals with mastery goals experience failure, i.e. achieve lower results than they had anticipated, they interpret the event as providing information regarding their effort in that particular situation, attribute failure to a lack of effort or ineffective strategy use and try to improve upon the experience (Dweck, 2000; Harackiewicz, Barron, Tauer, Carter, & Elliot, 2000). For performance-oriented students, feedback that indicates their failure may have detrimental consequences affecting their self-esteem, raising anxiety and leading to self-handicapping behavior (Ilgen & Davis, 2000; Ilies & Judge, 2005).

FIT can be quite effective in explaining the results of the above mentioned studies. Since performance-oriented students strive to outperform their peers, rather than focus on a task, negative feedback will be processed and interpreted through the prism of their self. In this case, information contained in a feedback message is unlikely to be used to improve performance and adjust strategies. Depletion of cognitive resources which results from students' focus on their selves and not the task will impede the constructive use of feedback and their performance is likely to decrease or stay the same. Conversely, mastery-oriented students, characterized by their focus on task and improvement of performance are likely to see feedback as a vehicle to enhance their skills and achieve better outcomes.

The paucity of studies investigating the effect of goal orientation on students' reaction to feedback calls for an inquiry that would shed light on the issue. The present research attempted to examine whether students' goal-orientation led to differential

patterns of response to feedback. A range of experimental conditions allowed for

obtaining a complex picture of the interactions among goal orientation, feedback source,

and the type of feedback message.

Self-efficacy

Bandura (1982b) referred to self-efficacy as individuals' confidence that they can

attain certain performance levels in specific areas of their lives. Academic self-efficacy

corresponds to one of the numerous domains of individuals' competences. It deals with

individuals' attitudes towards the learning process, their sense of direction and self-

expectations for academic performance. Self-efficacy has been shown to be both a

consequence and an antecedent of performance. As a consequence, beliefs about one's

competence on a task are influenced by individuals' prior outcomes and therefore, self-

efficacy is molded by their performance. As an antecedent, higher self-efficacy is

consistently shown to lead to greater use of diverse learning strategies, increased effort,

sustained persistence, and higher attainment on a variety of tasks (Bandura, 1997; Lee &

Klein, 2002; Linnenbrink & Pintrich, 2003; Schunk, 1990, 1995).

Students' level of self-efficacy may affect their reaction to feedback. Silver,

Mitchell, and Gist (1995) investigated the moderating role of self-efficacy in individuals'

responses to feedback. They found that high self-efficacy led to attribution of failure to

external and unstable causes, demonstrating that individuals with high sense of

competence are less likely to be discouraged by negative feedback and will make

protective attributions to preserve self-efficacy. Similarly, several studies demonstrate

that individuals high in self-efficacy are less likely to quit a task even after receiving

feedback indicative of their poor performance relative to those low in self-efficacy (Bandura & Locke, 2003; Zimmerman, Bandura, & Martinez-Pons, 1992).

Students' efficacy levels are often tied to the notion of goal orientation discussed above. Ames (1992) proposed that feedback indicating poor performance is interpreted differently depending upon students' goal orientation, especially for those with low self-efficacy. The damaging impact of negative performance feedback is found to be worse for those who have low self-competence beliefs when they operate under a performance goal orientation (Ilgen & Davis, 2000). Mastery goals may reduce pressure associated with trying to "look competent," free cognitive resources allocated to worrying, and therefore may be less damaging to future performance for individuals low in self-efficacy.

The results of a study conducted by Bouffard, Bouchard, Goulet, Denoncourt, and Couture (2005) agree with the aforementioned research as they showed interaction between self-efficacy and goal-orientation. In their experiment, students were assigned to mastery and performance goals conditions, and within each condition low and high self-efficacy were induced by presenting extremely negative or extremely positive feedback. Students' performance, use of strategies, and self-regulation ability were examined by administering self-report measures and conducting interviews. The researchers found that high self-efficacy students in the mastery goal condition exhibited more effective self-regulatory behaviors as compared to low self-efficacy students. Persistence differed between low and high self-efficacy groups within the mastery goal condition, as low efficacy students were more likely to quit than those high in self-efficacy. Additionally, differences in performance were found, with high efficacy students in the mastery

condition outperforming those low in self-efficacy. Interestingly, no evidence was found that students' self-efficacy had an effect on their task outcomes in the performance goal condition (Bouffard et al., 2005). The fact that the researchers were able to successfully manipulate self-efficacy by presenting students' with differential feedback indicates the importance of feedback as means for influencing students' affective, as well as cognitive aspects of performance.

Other researchers have investigated the effect of self-efficacy on feedback acceptance (Nease, Mudgett, & Quinones, 1999), joint effect of self-efficacy and descriptive feedback on students' performance (Thomas, 1993), and perceptions of feedback accuracy as dependent on individuals' self-efficacy levels (Jussim, Yen, & Aiello, 1995). The present study looked at the impact of different types of feedback on students' self-efficacy.

Summary

This chapter presented an overview of assessment literature in an attempt to elucidate the nature of feedback as a means for promoting students' learning. The field of assessment appears to have numerous contradictions and inconsistencies, and many researchers have stressed the need for carefully designed studies that would fill gaps in the current knowledge of the effects of feedback on students learning (Kluger & DeNisi, 1996). The present study inquired into students' reactions to feedback. By adopting a more inclusive view of how feedback contributes to individuals' performance and individual characteristics of motivation, affect, and self-efficacy, a broad perspective of the effects of different types of feedback on students learning was examined.

The following research questions guided the study:

- o What are students' reactions to differential feedback?

  - What are the effects of grades on students' performance?

  - What are the effects of praise on students' performance?

  - Does the source of feedback affect student' reactions to feedback?

  - Do grades, praise, and the source of feedback have differential effects on performance of students of different ability levels?

- o Do differential feedback messages affect students' personal characteristics of motivation, self-efficacy, and affect?

- o Do differential feedback messages affect students' perceptions of the accuracy and helpfulness of feedback?

- o Do students with different goal orientation respond to feedback differently?

CHAPTER III

Method

The present study was conducted using a sequential explanatory design (Creswell, 2003). It involved collection of quantitative data through the experiment, followed by qualitative data obtained through focus groups. The priority was given to the quantitative data, with the qualitative data being used to corroborate, refute, or augment findings from the experiment (Hanson, Creswell, Plano Clark, Petska, & Creswell, 2005).

The main component was a randomized experiment occurring within the context of an actual college course. The dependent measure included an authentic learning task with students working on an essay exam and then revising it based on feedback. The exam was a part of a course requirement and, therefore, was expected to be taken seriously by the participants. By conducting an experiment in a natural setting, ecological validity of the study was enhanced.

The experiment included three conditions with some students not receiving detailed feedback on their performance, other students receiving detailed feedback with an understanding that their feedback came from the course instructor, and a final group of students believing that their feedback was computer-generated. Additionally, the three conditions were crossed with two factors of grade (grade or no grade) and praise (praise or no praise), resulting in a 3 x 2 x 2 design.

The second part of the study consisted of focus groups with selected participants. The purpose of the focus groups was to help with interpretation and to increase understanding of the outcomes of the experiment. The present chapter describes participants, instrumentation, and procedures for both components of the study.

Participants

Experiment

Participants for the experiment were students enrolled in introduction to psychology courses at Rutgers University and The College of New Jersey taught by the same instructor. IRB approval to conduct the study was obtained from both institutions. One of the graded course assignments involved writing an essay on a relevant topic. Informed consent was obtained to use students' written answers for research purposes and to administer a series of questionnaires. Students who allowed the use of their response for research and completed several self-report questionnaires satisfied their general psychology research requirement. The sample size for the experiment was 464 students, with 409 students attending Rutgers University and 55 students attending The College of New Jersey. Separate analyses were run for Rutgers University and The College of New Jersey samples to compare the distributions of key variables included in the current study. Examination of the results showed that these variables were distributed normally for both samples, with nearly identical means and standard deviations. Therefore, the decision was made to merge Rutgers University and The College of New Jersey samples together, and treat them as one sample.

The participants ranged in age from 17 to 51, with a mean age of 18.9, and a standard deviation of 2.5. Two hundred and forty one (51.9%) participants were women and 223 (48.1%) men. The majority of the participants identified themselves as White (54.7%), 24.6% as Asian, 6.9% as Hispanic, 3.9% as Black, 6.0% as Other, and 3.4% chose not to respond. Of the 464 participants, 382 (82.3%) were born in the US, and 82

(17.7%) were not. Students also provided information about their native language. Three

hundred and seventy one students (80%) reported to be English-speakers, and 93 (20%)

to be native speakers of a language other than English. A summary of the personal

characteristics of the participants is provided in Table 1.

Table 1

Personal Characteristics of Study Participants

Variable	N	Percentage
College		
Rutgers University	409	88.1
The College of New Jersey	55	11.9
Gender		
Female	241	51.9
Male	223	48.1
Race/Ethnicity		
White	254	54.7
Asian	114	24.6
Hispanic	32	6.9
Black	18	3.9
Other	28	6.0
Did not respond	16	3.4
Country of Origin		
US	382	82.3
Other	82	17.7
Native Language		
English	371	80.0
Other	93	20.0

Focus Groups

Six focus groups, each consisting of eight to nine Rutgers University students, were conducted two weeks following the completion of the experiment. For each group, students were selected based on the source of feedback they received and the presence or absence of a grade. Thus, separate focus groups were held for students in the no feedback condition who received a grade, and those who did not receive a grade, for students in the instructor condition who received a grade and those who did not, and, finally, for students in the computer feedback condition with the grade and without the grade. Preliminary analyses of experimental data showed no significant effect of praise. Therefore, the decision was made to collapse the groups on the praise factor and include equal number of students who did and did not receive praise into each of the six groups. The composition of the six focus groups is presented in Table 2.

Table 2

Composition of the Focus Groups

№	Type of condition	Composition (students)	Total
1	Instructor feedback, grade	Praise (4)	8
		No praise (4)	
2	Instructor feedback, no grade	Praise (4)	9
		No praise (5)	
3	Computer feedback, grade	Praise (4)	8
		No praise (4)	
4	Computer feedback, no grade	Praise (4)	8
		No praise (4)	
5	No feedback, grade	Praise (4)	8
		No praise (4)	
6	No feedback, no grade	Praise (4)	8
		No praise (4)	

When students agreed to partake in the experiment, they were asked whether they would want to be a part of a focus group to share thoughts about their experience. Out of those who expressed their interest to be a part of discussion, four students from each experimental condition were randomly selected and assigned to the six focus groups. The students were advised that for participating, they would receive 5 points (out of a maximum of 100) towards their final exam score. The sample consisted of forty-nine students.

The participants ranged in age from 17 to 22, with a mean age of 18.8, and standard deviation of 1.3. Twenty-six participants were women (53%), and twenty-three were men (47%). Thirty participants identified themselves as White (61.2%), eleven as Asian (22.4%), five as Hispanic (10.2%), two as Black (4.1%), and one as Other (2%). Forty-one participants (83.7%) were US-born, and eight (16.3%) were born outside of the United States. Forty-three of the participants (87.8%) were native English speakers and six students (12.2%) reported to be native speakers of a language other than English. A summary of the personal characteristics of the participants is provided in Table 3.

Pilot

Prior to the administration of the main study, the study was piloted at Kean University with students enrolled in an equivalent introductory psychology course. IRB approval for the pilot study was obtained. Students participated in the pilot for 5 extra credit points towards their final grade. The sample size was comprised of forty students.

Instrumentation

The dependent measure for this research consisted of an essay examination, written during an initial 90-minute session and then revised after a one week interval. Additionally, the following instruments were administered: (1) The Learning and Performance Goal Orientation scale adapted from (Button, Mathieu, & Zajac, 1996); (2) The Posttest Index of Test Motivation (Wolf & Smith, 1995); (3) The Test Self-Efficacy Scale; (4) The Positive and Negative Affect Scale (Watson, Clark, & Tellegen, 1988); (5) a set of background questions.

Table 3

Personal Characteristics of the Focus Group Participants

Variable	N	Percentage
Gender		
Female	26	53
Male	23	47
Race/Ethnicity		
White	30	61.2
Asian	11	22.4
Hispanic	5	10.2
Black	2	4.1
Other	1	2.0
Country of Origin		
US	41	83.7
Other	8	16.3
Native Language		
English	43	87.8
Other	6	12.2

Performance Task

As a part of the course requirement, students were asked to write a 500-word expository essay demonstrating their understanding of theories of motivation discussed in class. The prompt for this assignment was chosen from the Educational Testing Service list of topics appropriate for first-year students. The original topic was:

> Sometimes we choose to do things that we do not really enjoy — jogging, studying, eating the right foods, and so on. Describe something you do by choice that you really do not enjoy. Explain why you continue to do it. Discuss the changes that might occur in your life if you were to stop this activity. (ETS, 2006, ¶ 10).

This topic was modified to incorporate a clear reference to theories of motivation. The ETS E-Rater team was contacted and their approval to adapt the topic was obtained. The resulting prompt was:

> Sometimes we choose to do things that we do not really enjoy — studying hard, eating the right foods, and so on. Describe something you do by choice that you really do not enjoy. Using theories of motivation, explain why you might continue to do it. Discuss the changes that might occur in your life if you were to stop this activity. Support your claims with specific examples from your life and the course reading.

Students were presented with an extensive rubric (see Appendix A) describing criteria for evaluation. The rubric was available during the task and could be consulted at any point in the writing or revising process. In order to make sure that students wrote essays of comparable length, an indicator displayed a real-time word count.

Goal Orientation Measure

Button et al.'s (1996) Learning and Performance Goal Orientation scale was used to assess students' goal orientation. This measure consists of two scales: (1) a ten-item

performance orientation scale, and (2) a ten-item learning orientation scale, for a total of 20 items (see Appendix C for the list of items).

For the Performance Goal Orientation, answers are based on a 7-point scale that ranges from (1) "Strongly Disagree" to (7) "Strongly Agree." Agreement with these items suggests strong desire to obtain favorable judgments of one's competencies or, conversely, a desire to avoid negative judgments of one's competencies. Disagreement indicates little concern for performing better than others or making errors.

Learning goal orientation items are also rated on a 7-point scale that ranged from (1) "Strongly Disagree" to (7) "Strongly Agree." Agreement with these items shows a strong desire to learn new skills, master complex assignments, and develop alternative strategies when working on a difficult task. Disagreement suggests little concern for mastering tasks or gaining competency.

The coefficients alpha reported by the authors were .76 and .82 for the Performance Goal Orientation Scale, and was .79 and .71 for the Learning Goal Orientation Scale (Button et al., 1996). Coefficients alpha reported by Ward, Rogers, Byrne, and Masterson (2004) were .71 and .79 for Performance, and .82 and .87 for the Learning Goals Scale.

Button et al. (1996) maintain that the two dimensions of goal orientation constitute two separate theoretical entities and are uncorrelated. This assumption was tested in four separate studies by conducting confirmatory factor analyses and calculating correlation coefficients (Button et al., 1996). Strong evidence was provided showing that the two-factor model (as opposed to one) yielded a significantly better fit. The results also indicated that the two dimensions were not related. Therefore, learning goals and

performance goals are neither mutually exclusive, nor contradictory. It is possible for an individual to simultaneously strive to master one's skills and to perform better than others. Whereas some individuals may favor one type of goal over the other (i.e., be predominately learning goal or performance goal oriented), other combinations are also plausible. Some students may be motivated by both types of goal, and others may be generally apathetic and will be equally disinterested in learning and performance goals (Button et al., 1996; Mattern, 2005). For the purposes of the present study two separate scale scores were computed and further recoded into one variable indicating one of the four types of goal orientation (i.e., performance, learning, both, or neither).

Test Motivation Measure

The Posttest Index of Test Motivation (Wolf & Smith, 1995) consists of eight 7-point Likert-type items bounded by "Strongly disagree" and "Strongly agree." This scale has been successfully used to test how motivated students were to do well on a task in question (Spencer, 2005; Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995). High scores on the scale indicate that students had a strong desire to do well on the exam they just took and exerted all the necessary effort to ensure success. Lower scores suggest a lack of interest in the process or the outcome of the exam. See Appendix D for the list of items.

The reliability coefficients reported in the literature are .89 (Spencer, 2005) and .87 (Wolf et al., 1995), which are comparable to the alpha coefficients of .84 to .87 of the original scale reported by Wolf (1993).

Test Self-Efficacy Measure

The Test Self-Efficacy Scale was constructed as part of a dissertation research study conducted by Spencer (2005). The measure consists of eight Likert-type items. The answers were based on a 7-point response scale ranging from (1) "Strongly Disagree" to (7) "Strongly Agree." The items are treated as an additive scale, resulting in a total test self-efficacy score for each person. Higher scores on the measure indicate students' confidence in their performance on the test, and lower scores suggest doubt in their ability to have done well on the task in question. The reported alpha coefficient of the instrument is .86 (Spencer, 2005). The list of items is presented in Appendix E.

Measure of Affect

The Positive and Negative Affect Scale (PANAS) is a 20-item self-report measure of positive and negative affect developed by Watson et al. (1988). The PANAS is believed to provide independent measures of positive and negative affect. In the present study the scale was accompanied with momentary instructions for measuring students' current affective state (see Appendix F). The participants were asked to indicate the extent to which they experienced the affective states described by the PANAS adjectives on a 5-point scale ranging from "Slightly/Not at all" to "Extremely."

Since its development, the measure has been widely used in research for diverse purposes. Its popularity may be attributed to its brevity and solid psychometric characteristics. The scales were shown to be highly internally consistent, largely uncorrelated, and stable (Watson et al., 1988). Relative independence of the two scales may seem to be counter-intuitive, as happiness and sadness represent the opposite poles

of the mood continuum. Watson, Wiese, Vaidya, and Tellegen (1999), however, argue

that the dimensions of PA and NA represent the subjective components of the bio-

behavioral systems of approach, known as the behavioral inhibition system (BIS), and

withdrawal, referred to as the behavioral engagement system (BES) (Crawford & Henry,

2004; Watson et al., 1999). These adaptive systems are viewed as separate, but not

entirely independent of one another, thus accounting for the moderate correlations

reported between the constructs of PA and NA (Watson et al., 1999). In the present study,

two additive indices were computed, resulting in separate PA and NA scores for each

participant.

The reported alpha coefficients of the positive affect scale range from .86 to .95,

and the negative affect scale from .84 to .92 (Crawford & Henry, 2004; Ilies & Judge,

2005; Jolly, Dyck, Kramer, & Wherry, 1994; Roesch, 1998).

Demographic Data

A short demographic questionnaire was administered to the research participants.

The participants were asked to report their age, gender, race, native language, and

country of origin.

The list of instruments administered and time of their administration is presented

in Table 4.

Table 4

Instrumentation and Time of Administration

Instrument	Measures	Time of administration	Location
I	Demographic questionnaire (7 items)	First session of the experiment; before students begin the exam	Appendix B
II	Goal orientation questionnaire (20 items)	First session of the experiment; before students begin the exam	Appendix C
III	Essay exam	First session	
IV	Positive affect and negative affect scale (18 adjectives)	Second session; after feedback was presented, but before students begin revising	Appendix F
V	Post-test index of test motivation (8 items)	Second session; after the revised essay was submitted	Appendix D
VI	Post-test self-efficacy (8 items)	Second session; after the revised essay was submitted	Appendix E
VII	Accuracy of feedback (1 question)	Second session; after the revised essay was submitted	
VIII	Helpfulness of feedback (1 question)	Second session; after the revised essay was submitted	

Procedures

The experiment involved computer administration and was conducted on two sessions separated by one week. A custom data collection program and an interactive Web site had been created to satisfy specific requirements of this study.

First Session

All students enrolled into the two introductory psychology courses were scheduled to come to a computer lab to take their exam. The course instructor discussed the specifics of the exam and answered students' questions. The experimenter collected consent forms from students who agreed to take part in the study. By agreeing to take

part in the study students consented to allow their responses to be used for research

purposes and to answer a series of demographic questions, as well as questions

concerning their motivation, affect, goal orientation, and self-efficacy.

All students logged into the dedicated website and were assigned a unique code

derived from their names. This code was used as identification in the experiment, and a

cross-reference table of codes and students' names was available to the course instructor

to allow for grading. Students who chose not to participate in the research study were

encouraged to begin their work on the exam. Students who consented to participate in the

research were asked to complete the Goal Orientation Questionnaire and a short

demographic questionnaire.

For the main task of the experiment, students were presented with the instructions,

the grading rubric, and were then asked to begin their essay. Latency measures such as

delay before commencing, number of referrals to the grading rubric, and time spent

writing were recorded. Students submitted their work which was saved in the system, and

were then thanked for their performance and were reminded to come back to the

computer lab in one week for the second part of the study. The layout of the essay-

writing screen is presented in *Figure 1*.

The following criteria will be used to evaluate your work (hover the mouse to read)

Sometimes we choose to do things that we do not really enjoy - studying hard, eating the right foods, and so on. Describe something you do by choice that you really do not enjoy. Using theories of motivation, explain why you might continue to do it. Support your claims with specific examples from your life and the course reading.

I would like to write about

Word count: 6

Save and continue editing

Go back   Submit

Figure 1. Layout of the essay-writing screen during the first session.

Scoring

ETS allowed the use of their proprietary software package Criterion in the present research. Criterion is comprised of two modules: E-rater and Critique. E-rater is a scoring application that extracts linguistically based features from an essay and uses a statistical model of how these features are related to overall writing quality in order to assign a holistic score to the essay. Parallel to E-rater, the Critique component of Criterion assesses and provides feedback for errors in grammar, usage, and mechanics, identifies the essay's structure, recognizes undesirable stylistic features, and provides diagnostic annotations within each essay (Attali, 2004). The principal purpose of the E-rater and

Critique software is to automatically score students' essays and generate itemized

feedback, whereas Criterion is an interactive online application that provides access to E-

Rater and Critique (Burstein, Chodorow, & Leacock, 2004).

Several requirements for the administration of the experiment did not allow the

use of off-the-shelf software in conducting the present study. Those include the non-

standard nature of the task, repeated log-ins by the same participant at different points in

time, differential feedback, collection of latency measures, and the combination of

feedback from the computer (supplied by E-Rater) and humans (course instructor and

experimenter). Hence, a custom Web site was created to satisfy the requirements, and a

software program was written to support this Web site. This Web site accessed E-Rater

and Critique directly, so the Criterion interface was not used. Access to the Web site was

restricted to study administrators, course instructors, and participants.

The total exam score presented to the students was comprised of two separate

components: the E-rater score (ranging from 0 to 6) and the content score provided by the

instructor and the experimenter (ranging from 0 to 6, including half-points). The final

score was calculated as a weighted average of the two scores and converted into a scale

of 100. The E-rater score contributed 30% to the total score, and the content score

contributed 70% to the total score.

The E-Rater software package had been trained to rate essays written on the

prompt selected for the present study. Students' essays were scored on all of the

aforementioned characteristics including mechanics, grammar, spelling, and stylistic

features, and a holistic score was assigned to every student. For several experimental

conditions, the feedback provided by E-Rater was modified to satisfy the requirements of

specific feedback conditions described below. A portion of the detailed feedback screen

is presented in *Figure 2*.

A drive can be described as an internal force. However, there are many internal and external forces that motivate people. There are many things people do just because they love doing it, so they continue the behavior **becuase** #7 it makes them feel good. People are also motivated by external rewards such as the title of "First Place" or even things like social acceptance. The problem with this type of motivation is that people will stop enjoying what they are doing **becuase** #8 they are only doing it for the reward. #70 I enjoy learning new things, but most of the time I study and work hard just to get a good grade. I find that school ~~...~~

#70: Organization and Development: You may need to fully explain extrinsic vs intrinsic motivation. Use theory-specific terminology when presenting your argument.

question of "what can get me a good ~~...~~ Another thing I noticed in my school work is that I do not always try my best because I can then use that as an excuse if I do poorly. This is referred to as a "Fear of Failure." If I tired hard on something and **recieve** #9 a bad grade, it would make me feel terrible; but if I do not try hard, receiving a bad grade is easier to deal with. #4 There are many theories and concepts that deal with motivation, and one of the questions that come up is "why do we do things that we #78 do not want to do?" Going to school is a big obligation in my life right now, but hopefully understanding my motives can help me find the best way to go through it.

Figure 2. Detailed feedback screen with a popup message for a specific feedback item.

Additionally, two raters (the course instructor and the experimenter) ensured that

the content was covered properly. Prior to scoring the main experiment, a series of

calibration sessions were held to ensure inter-rater reliability between the two raters. A

detailed rubric was developed which provided criteria for evaluating the content of

students' essays (see Appendix G). The inter-rater reliability was .96 for the first exam

score and .98 for the second exam score. In case of a discrepancy in ratings, the average

of the two raters' scores was taken. There were no differences in ratings larger than one

point, which is indicative of the high level of calibration between the two raters. The

instructor and the experimenter were oblivious to the students' identities, i.e. blind

scoring was employed. To provide feedback on the content of students' essays, several

standard comments were written. These comments were modified slightly depending on

the experimental condition, so that some comments sounded as if they came from a

computer and others from the professor.

After all of the initial essays were scored, blocking was used to assign participants to three experimental conditions so that the resulting groups had equivalent numbers of students with high, medium, and low scores.

Each student was assigned to one of the three feedback conditions:

1. *No feedback condition*. This group received no detailed feedback.

2. *Instructor feedback condition*. This group received a combination of E-Rater-generated feedback regarding mechanics and style, and content-related comments and suggestions, with the understanding that all the comments were generated by the course instructor. All comments were written in a reserved neutral fashion, but in way that was clear that they came from a person rather than a computer. Also, students were addressed by their first name. To make sure that the source of feedback was clear to the participants, a clip-art picture of a typical college professor was displayed in the corner of every exam screen, and the following instruction were provided: "During this session, you will be able to edit and improve the essay you wrote the first time, based on detailed feedback I have given you on content, grammar, punctuation, spelling, sentence structure, and the overall quality of your essay. PLEASE READ MY COMMENTS CAREFULLY and do your best to use them — it should really help you get a better score."

3. *Computer feedback condition*. Students in this group received feedback equivalent to the one in the previous condition with the understanding that all the comments were generated by the computer. The following instructions were provided: "During this session, you will be able to edit and improve the essay you wrote the first time, based on detailed feedback generated by an intelligent computer system designed to read and critique essays. The computer will give you feedback on content, grammar,

punctuation, spelling, sentence structure, and the overall quality of your essay. PLEASE READ THE COMPUTER'S COMMENTS CAREFULLY and do your best to use them — it should really help you get a better score."

A picture of the computer was displayed on every screen. E-rater comments were taken in their original form, and the additional comments concerning the content and adequacy of the use of course-related constructs matched the style of the computer comments, and were impersonal and neutral. Students were not referred to by their first names. A comparative table of the comments received by students in the computer and instructor conditions is presented in Table 5. The number of words contained in the feedback messages was controlled for each participant to make sure that all students received feedback of similar length. This allowed for estimation of the time students spent reading feedback, avoiding the message-length confound.

Table 5

Comparison of Comments Received by Students in the Instructor and Computer Conditions

Type of comment	Instructor	Computer
Mechanics	*Name*, please break your essay into paragraphs so I can see the structure.	Please break your essay into paragraphs so that the structure can be detected.
	Name, this sentence is a fragment. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.	This sentence may be a fragment. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.
	Name, these sentences begin with coordinating conjunctions. Try to combine the sentence that begins with *but* with the sentence that comes before it.	These sentences begin with coordinating conjunctions. A sentence that begins with *and, but*, and *or* can sometimes be combined with the sentence that comes before it.
Content	*Name*, a good essay usually contains three main ideas, each developed in a paragraph. Use examples, explanations, and details to support and extend your main ideas. Try to center them around the theories of motivation I discussed in class. Include details and theory-specific terminology.	A good essay usually contains three main ideas, each developed in a paragraph. Use examples, explanations, and details to support and extend your main ideas. Center them around the theories of motivation. Include details and theory-specific terminology.
	Name, please discuss all of the components of the Drive reduction theory: need, drive, action, and homeostasis. You are missing two of the components.	You may need to discuss all of the components of the Drive reduction theory: need, drive, action, and homeostasis.
	Name, discuss all of the components of Atkinson's theory: expectancy, value and the need for achievement. You are missing one of the components.	Discuss all of the components of Atkinson's theory: expectancy, value and the need for achievement. You may be missing some of the components.

Additionally, the three conditions were crossed with two factors of grade (grade/no grade) and praise (praise/no praise) resulting in a 3x2x2 experimental design. The groups formed by the factor crossings are presented in Table G1.

Praise was presented in the form of a standard comment preceding the rest of the feedback. There were three levels of praise which differed depending on the grade students received for their original essay. Students in the instructor feedback condition were referred by their first name, whereas students in both the computer feedback and no feedback conditions were not addressed by their first name. See Table 6 for the three levels of praise for each of the three feedback conditions.

Table 6

Levels of Praise for the Instructor, Computer and No Feedback Conditions

Exam Score	Instructor Feedback	Computer Feedback	No Feedback
80 to 100	*Name*, you made an excellent start with this essay! I still see room for improvement, so take some time and make it really great.	You made an excellent start with this essay. The data indicate there is still room for improvement, so take some time to make it better.	You made an excellent start with this essay! There is still room for improvement, so take some time and make it really great.
70 to 79	*Name*, you made a very good start with this essay! I still see room for improvement, so take some time and make it really great.	You made a very good start with this essay. The data indicate there is still room for improvement, so take some time to make it better.	You made a very good start with this essay! There is still room for improvement, so take some time and make it really great.
69 and below	*Name*, you made a good start with this essay! I still see room for improvement, so take some time and make it really great.	You made a good start with this essay. The data indicate there is still room for improvement, so take some time to make it better.	You made a good start with this essay! There is still room for improvement, so take some time and make it really great.

Second session

The participants were asked to return to the computer lab in one week. They logged into the system and were shown their graded essay with its corresponding feedback. The amount of time they spent reading the feedback was recorded. Prior to moving to the essay revision screen, students were asked to fill out the Positive and Negative Affect scale. The participants were then prompted to make revisions and resubmit their essay based on the feedback they received. Students could refer to the grading rubric and to their feedback comments at any point of the session by hovering their mouse over hotspots in the feedback text.

Students in the control group were encouraged to reread their essays, consult the rubric, and work on improving their work. After the participants submitted their revised essays, they were asked to make a judgment concerning the accuracy and helpfulness of the feedback. They were also asked to complete the Post-Test Index of Test Motivation, and the Test Self-Efficacy scale.

Focus Groups

After the preliminary analyses were conducted, eight students were randomly selected from each condition for participation in one of the six focus groups (see Table 2 for description). Assuming that the participants would be more candid in the presence of a person they already know, the experimenter served as a focus group moderator. One student who was not selected to participate volunteered to join the discussion. The

decision was made not to exclude her contribution from the analysis. Therefore, all the data were retained.

Each focus group took place in a lounge area of one of the university buildings. Students were provided with an information sheet informing them that their participation was voluntary, they were under no obligation to answer any questions once in the focus groups, and their responses could not be traced back to them in the ensuing manuscript or publication. Students were informed that they were going to be audio taped, and that the experimenter would be the only person with access to the tapes and transcripts. Once transcribed and analyzed, with all the identifiers removed, the transcripts could be shared with parties directly involved in the study.

A total of six focus groups were conducted, each approximately 40 minutes long. A semi-structured discussion was used to elicit participants' responses to the set of questions. It employed a careful specification of topics to be covered before the discussion, but flexibility in terms of the sequence and the phrasing of the questions in the course of the focus group. The experimenter followed the flow of conversation, changing the order of questions when it made sense to do so. The conversational format of a semi-structured interview is usually less intimidating than a formal predetermined set of questions (Patton, 1990). Topics for discussion with the participants were closely tied to the research questions. The semi-structured discussion guide used in each of the six focus groups had the following questions:

- How did you react to the feedback? How did you go about revisions?
- What did you feel when you received your feedback? What was the best and the worst part of the feedback you received?

- Did you trust your feedback? Did you find it accurate? Did you think it was fair and worthwhile?

- How did your grade help you? Would you have liked to know your grade before you began your revisions?

- Was the praise helpful? Would you have liked to receive praise on your performance?

- What would be your ideal kind of feedback?

- Did you think getting a chance to revise was good?

Therefore, the discussions focused on understanding students' reactions to feedback, their affective and cognitive responses to it, and on the appropriateness of feedback for the task. The students were asked to discuss what they believed was the most helpful and the most discouraging feature of the feedback they received, and encouraged to think about the optimal type of feedback they wish they could have received for the task.

Data Analysis

The sequential design of the study with the primary focus on the quantitative data required that the experimental data were analyzed first, with the analysis of the focus group data to follow. Quantitative data were analyzed using SPSS 13.0 statistical software. The statistical analyses were selected to yield the output needed to address the main research questions. Descriptive statistics were calculated to describe the sample and to analyze students' responses to the instruments administered in the course of the study. Analysis of covariance (ANCOVA) was used to address the research question that

inquired into the difference in performance based on the feedback that students received. Descriptive statistics were used to separate students into groups based on their writing ability, and differences in performance as dependent on the type of feedback were analyzed for each of the ability group using ANCOVA. To test whether students' reported motivation, self-efficacy, and affect varied depending on the type of feedback they received, multivariate analysis of variance (MANOVA) was employed. Differences in judgments about helpfulness and accuracy of feedback were analyzed using MANOVA. To investigate whether differential feedback affected students' performance differently depending on their goal orientation, students were split into four groups based on their goal orientation, and ANCOVA analysis was conducted.

After the experimental data were analyzed, the analysis of the focus group data was conducted with the aim to support, refute, or expand upon the findings of the experiment. Hence, the analysis was closely tied to the main findings of the experiment. It was not the intent to read the data inductively, searching for concepts and categories that were not directly connected or had no relevance to the research questions of the present study and the main findings of the experiment.

All focus group discussions were transcribed verbatim. Formal analysis of the data records began with the use of deductive coding. The data were read and coded according to predetermined categories drawn from the research questions and the main findings of the experiment. Categories included students' interpretations of the effects of grades, praise, and feedback source on their performance, their perceptions of the effect of feedback on motivation, affect, and self-efficacy, and their views on what would be ideal feedback. The coding was carried out separately for each of the six focus groups.

Summaries of findings were compiled for each group to present a general picture of participants' experiences and their reactions to the specific type of feedback they received. The data were then compared across the six groups, with the analysis being guided by the findings of the experiment. This step allowed for determining similarities and differences for participants who had different experience in terms of feedback, and, most importantly, it brought together the two components of the present study.

CHAPTER IV

Results

The results are divided into seven sections. First, descriptive information

regarding the main variables and instruments in the study is presented. Second,

differences in responses to feedback messages are examined for the entire sample. Third,

differences in students' performance are analyzed and presented for students of different

writing ability levels. Fourth, differences in motivation, self-efficacy, and affect are

explored for all feedback conditions. Fifth, the perceived helpfulness and accuracy of

feedback is analyzed. Sixth, students' responses to feedback as dependent on their goal

orientation are investigated. Finally, a summary of findings is presented for each of the

six groups followed by a description of the results across the groups.

Descriptive Statistics

Means, standard deviations, and intercorrelations of all major variables in the

study are presented for purposes of reference in Table 7. No univariate or multivariate

outliers were identified upon the examination of the data. Thus, all participants were

retained in the analysis of the hypotheses of the study.

A series of coefficients alpha were calculated to examine the reliability of the

scales administered in the course of the study. All measures demonstrated sufficient

reliability, with alphas ranging from .80 to .89. Every item was found to contribute to the

internal consistency of the corresponding scale, and removal of any item would result in

the decrease of the scale's reliability coefficient. Alpha coefficients of the key measures

are presented in Table 8.

Table 7

Descriptive Statistics and Intercorrelations of Study Variables

№		M	SD	1	2	3	4	5	6	7	8	9
1	Exam 1 Grade	74.42	8.28	—								
2	Exam 2 Grade (revised essay)	78.94	8.72	.74***	—							
3	Performance Goal Orientation Score	50.17	8.40	.04	.01	—						
4	Mastery Goal Orientation Scale	57.25	7.97	-.02	-.01	.12	—					
5	Positive Affect Scale	29.86	7.17	.02	-.02	-.04	.13**	—				
6	Negative Affect Scale	24.00	7.51	-.14**	-.06	.15**	-.04	-.06	—			
7	Posttest Index of Test Motivation	48.19	6.79	.09*	.11*	.02	.16**	.30***	.08	—		
8	Test Self-Efficacy Scale	44.44	6.77	.24***	.23***	.00	.23***	.29***	-.22***	.37***	—	
9	How accurate was the feedback?	4.85	1.88	-.01	.19***	-.02	.02	.09	.11*	.01	-.09	—
10	How helpful was the feedback?	4.75	2.06	-.04	.21***	.01	.04	.11*	.12**	.11*	-.05	.84***

Note. For the Self Efficacy and Positive Affect Scales N = 462. For the remaining measures N = 463.
* p < .05, ** p < .01, *** p < .001.

Table 8

Reliability Coefficients for the Study Instruments

Name	Number of items	Number of valid responses	Coefficient α
Performance Goal Orientation Scale	10	462	.80
Learning Goal Orientation Scale	10	463	.87
Posttest Index of Test Motivation	8	463	.85
Test Self-Efficacy Scale	8	462	.86
Positive Affect Scale	10	463	.89
Negative Affect Scale	10	463	.86

Analyses of the Effects of Treatments on the Final Exam Score

The main hypothesis of the study stated that students' final performance on the essay exam would vary depending on the type of feedback they received on the draft version of their work. A 3 x 2 x 2 analysis of covariance (ANCOVA) with the source of feedback (x 3), grade (x 2), and praise (x 2) conditions as factors and the grade for the first exam (before revisions) as a covariate, examined differences in the final grades for the essay exam. The Bonferroni adjustment was employed to control for Type 1 error.

Looking first at the interaction effects, ANCOVA revealed a significant disordinal interaction between grade and praise, $F (1, 450) = 6.00$, $p < .05$, $\eta^2 = .04$. *Figure 3* shows that under the grade condition, scores were higher when praise was presented, M = 79.26, SD = 5.12, than when praise was not presented, M = 77.69, SD = 5.12. For the no grade condition, scores were higher when praise was not presented, than when praise was presented (M = 79.82, SD = 5.12, for the praise condition, and M = 79.06, SD = 5.13, for the no praise condition). Means and standard deviations are presented in Table 9. This finding is addressed again after the main effects have been presented.

Table 9

Estimated Marginal Means and Standard Deviations of the Final Exam Score by Grade

and Praise

		M	SD	N
No grade	No praise	79.82	5.12	118
	Praise	79.06	5.13	115
Grade	No praise	77.69	5.12	115
	Praise	79.26	5.12	115

Note. Adjusted means after controlling for the first exam score.
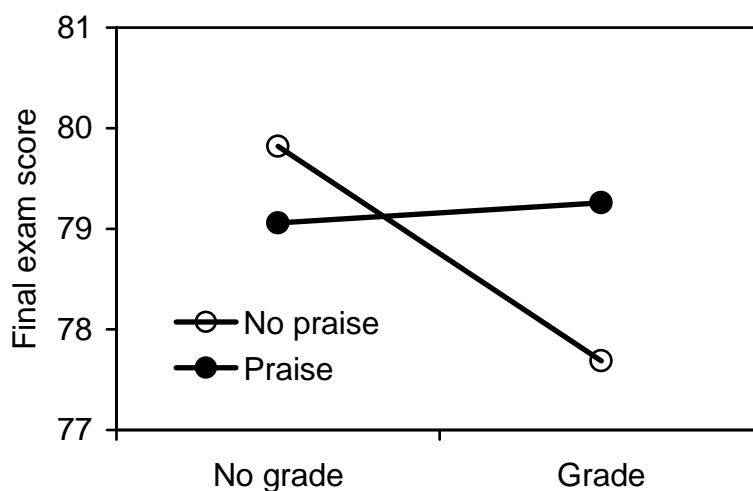
Figure 3. Mean final exam score as function of grade and praise.

The interaction between grade and praise is statistically significant, but the effect

size was found to be quite small. A potential explanation for such occurrence may lay in

the nature of the analysis. The analysis of covariance included a powerful covariate (the

initial test score), which accounted for a large proportion of the variance in the dependent

variable. As a result, less variation was left to be explained by the experimental

treatments and their interactions. Additionally, the small effect size can be attributed to the relatively large sample size that can make small effects significant.

There was also a significant interaction between grade and feedback source, F (2, 450) = 5.54, $p < .01$, $\eta^2 = .08$, see *Figure 4*. In the no feedback condition, scores were higher for students who received a grade, M = 75.37, SD = 5.12, as compared to those who did not receive a grade, M = 74.65, SD = 5.12. Under the instructor condition the opposite trend was observed. Students' final exam scores were relatively high when grade was not presented (M = 82.74, SD = 5.13), but they were lower for students to whom their grade was presented (M = 79.63, SD = 5.12). Under the computer condition, students' scores remained almost the same, slightly lower for those who received the grade (M = 80.93, SD = 5.12, for the no grade condition, to M = 80.44, SD = 5.12, for the grade condition). Means and standard deviations are presented in Table 10.

Table 10

Estimated Marginal Means and Standard Deviations of the Final Exam Score by Grade and Source of Feedback

		M	SD	N
No grade	No feedback	74.65	5.12	80
	Computer	80.93	5.12	79
	Instructor	82.74	5.13	74
Grade	No feedback	75.37	5.12	75
	Computer	80.43	5.12	80
	Instructor	79.63	5.12	75

Note. Adjusted means after controlling for the first exam score.
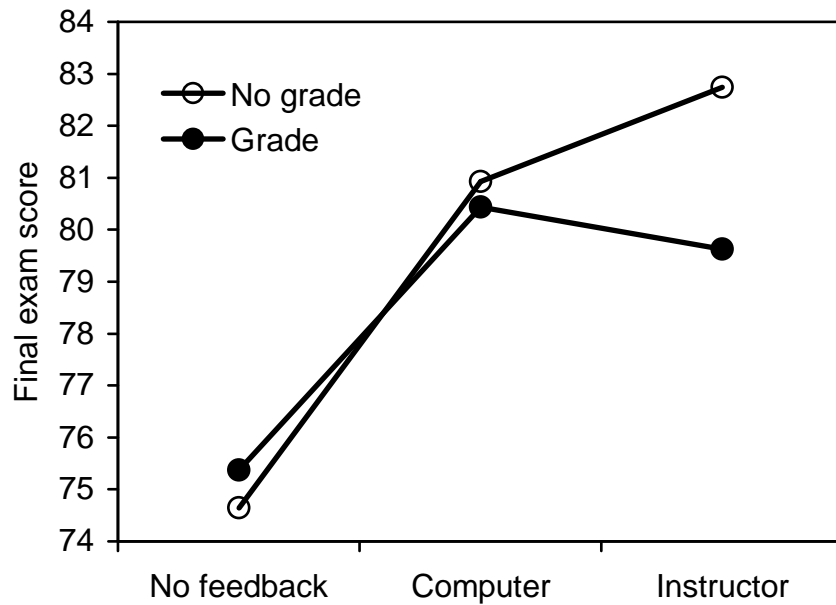
Figure 4. Mean final exam score as function of grade and feedback source.

Additionally, there was a strong significant main effect of the source of feedback on students' final grade, $F(2, 450) = 69.23$, $p < .001$, $\eta^2 = .24$, with students who did not receive detailed feedback demonstrating lower performance than those who received detailed feedback from either the computer or the instructor (see Table 11 for means). Post-hoc analysis revealed no differences in students' performance between computer and instructor conditions. There was also a significant difference in the final exam grade between students in the grade condition and those in the no grade condition, $F(1, 450) = 4.07$, $p < .05$, $\eta^2 = .04$. Students who were shown the grade they received for their first draft performed less well on the final version than those who were not shown their grade; see Table 11 for corresponding means.

Table 11

Means and Standard Deviations of the Final Exam Scores by Source of Feedback, Grade, and Praise

		No Grade			Grade			Total		
		No Praise	Praise	Total	No Praise	Praise	Total	No Praise	Praise	Total
No Feedback	M	73.80	74.38	74.09	75.11	76.24	75.67	74.44	75.27	74.85
	SD	8.57	9.21	8.84	8.56	7.60	8.07	8.54	8.47	8.49
	N	40	40	80	38	37	75	78	77	155
Computer	M	81.15	79.75	80.44	79.80	80.28	80.04	80.47	80.01	80.24
	SD	8.43	8.97	8.68	7.07	8.36	7.70	7.75	8.62	8.18
	N	39	40	79	40	40	80	79	80	159
Instructor	M	83.85	83.26	83.57	78.41	81.74	80.09	81.20	82.47	81.82
	SD	7.60	7.56	7.53	7.84	7.92	8.01	8.14	7.74	7.94
	N	39	35	74	37	38	75	76	73	149
Total	M	79.55	78.95	79.25	76.80	79.16	78.63	78.69	79.20	78.94
	SD	9.20	9.32	9.24	8.02	8.24	8.15	8.66	8.78	8.71
	N	118	115	233	115	115	230	233	230	463

No significant effects were found for the praise condition, or for interactions between praise and source of feedback, and among praise, grade, and source of feedback, $F (1, 450) = .71, p > .05, F (1, 450) = .28, p > .05, F (2, 450) = 1.66, p > .05$, respectively.

General Description of the Final Exam Score Variable as a Function of Experimental Factors

For the sake of the facility of data interpretation, praise and grade variables were recoded into a new combined variable with four levels defined as neither, grade only, praise only, and both. A line chart was used to summarize values of the new variable within the three conditions of the feedback source, using the means of the outcome variable, final exam grade (see *Figure 5*). Note that of all the trends discussed below,

main effects for grade and feedback source, as well as interactions between grade and

praise, and grade and feedback source, were found to be statistically significant.
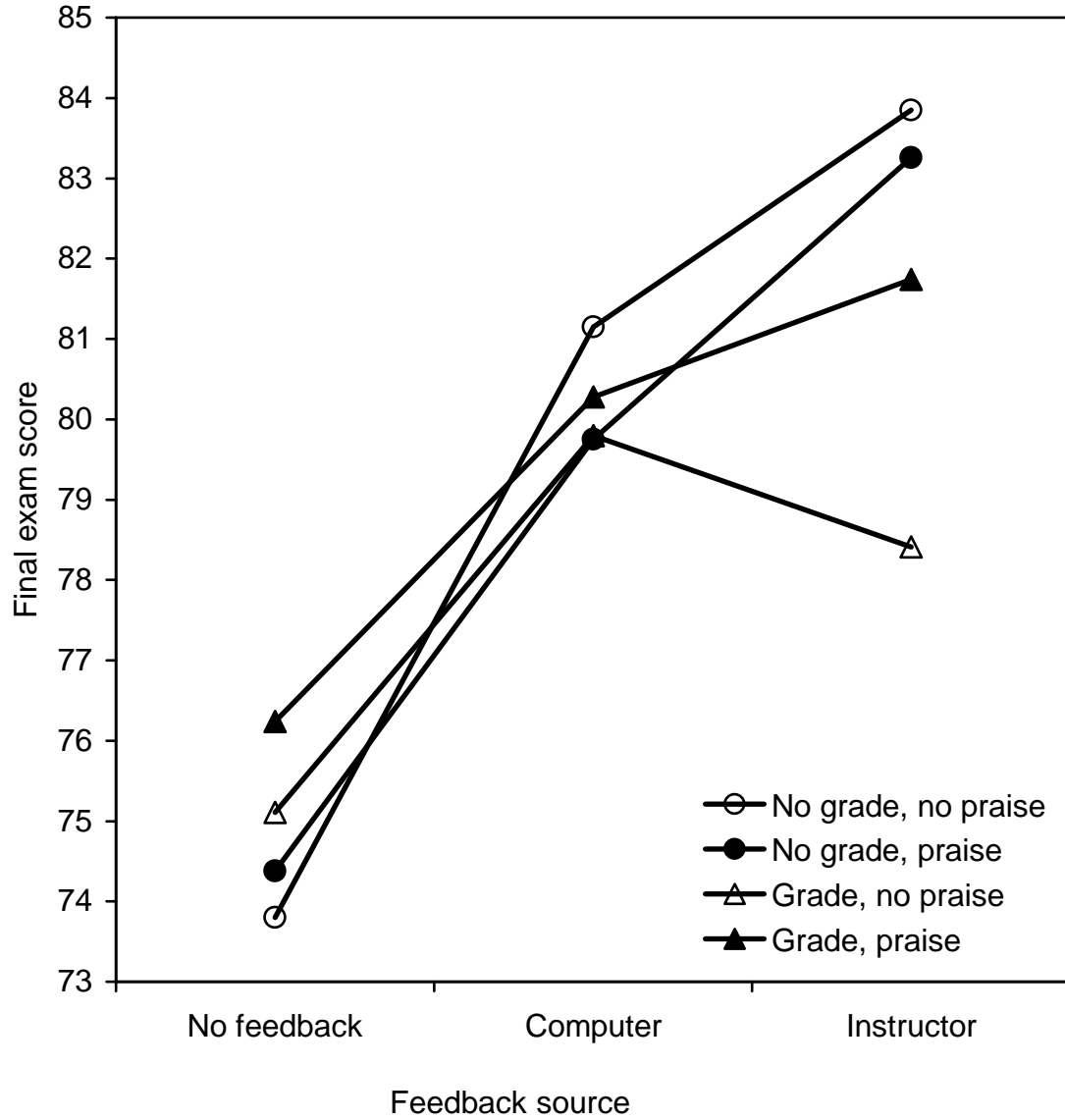
Figure 5. Mean final exam score as function of grade, praise, and feedback source.

Figure 5 seems to reveal several fundamental ideas. First, the importance of receiving detailed feedback in general is clear. Feedback is related to higher performance whether it is perceived to come from the computer or the instructor. Second, the presence of a grade appears to depress performance when the grade is perceived to come from the instructor. Third, the presence of praise from the instructor appears to ameliorate the negative impact of a grade. Fourth, if no detailed comments are provided, grades and praise lead to improved performance. In other words, any kind of feedback is better than no feedback at all. These results seem to indicate that the most effective form of feedback in this setting is descriptive feedback from the instructor not accompanied by grades or praise. Obviously, students need feedback to improve their performance, and the best kind of feedback appears to be impartial, neutral information without emotionally laden components such as grades and praise.

Analysis of Differences in the Final Exam Score for Students of Different Writing Abilities

To answer the research question concerning the effects of grade, praise, and the source of feedback on the performance of students of different ability levels, the following steps were taken. A frequency analysis was run for the first exam score. The analysis revealed a mean of 74.42, SD = 8.28, and a range from 50 to 96 for the final exam score. The analysis of frequency tables showed that 25% of the sample scored at or below 69 (equivalent to letter grades D and F), about 50% received a score between 70 and 79 (equivalent to the letter grade C), and the remaining 25% obtained a score at or above 80 (equivalent to letter grades B and A). Based on these cut points, students were

identified as being of low (N = 116), medium (N = 217), and high (N = 130) ability

levels.

The dataset was then split on the ability level grouping variable, and a series of 3

x 2 x 2 ANCOVAs were run with the source of feedback (x 3), grade (x 2), and praise (x

2) as factors, and the first exam grade as a covariate. These analyses examined

differences in the final exam scores for students in each ability group. Pairwise

comparisons were performed between each pair of the feedback source when ANCOVA

was found to be significant. The Bonferroni adjustment was employed to control for Type

1 error.

Low ability students

For the low ability students, the analysis revealed a significant grade by feedback

source interaction, F (2, 103) = 5.27, $p < .01$, $\eta^2 = .10$; see *Figure 6*. In the no feedback

condition, scores were higher for students' who received a grade, M = 67.85, SD = 6.64,

as compared to those who did not receive a grade, M = 64.15, SD = 6.75. As shown in

Figure 6, the overall scores were relatively low for this group. Under the instructor

condition, students' final exam scores were relatively high for the no grade condition, but

they were lower when the grade was presented (M = 77.24, SD = 6.86, for the no grade

condition, M = 72.07, SD = 6.65, for the grade condition). Under the computer condition,

students' scores were higher when the grade was presented (M = 72.07, SD = 6.64, for

the no grade condition, M = 75.50, SD = 6.71, for the grade condition). Means and

standard deviations are presented in Table 12.

There was also a significant effect for the source of feedback, F (2, 103) = 18.78,

$p < .001$, $\eta^2 = .28$, with students in the control condition who received no feedback

scoring significantly lower than those in either the instructor ($p < .01$) or computer

conditions ($p < .01$). See Table 13 for means and standard deviations. No differences

were revealed between the computer and instructor conditions ($p > .05$).

Table 12

Estimated Marginal Means and Standard Deviations of the Final Exam Score by Grade

and Source of Feedback for Low Ability Students

		M	SD	N
No grade	No feedback	64.15	6.75	19
	Computer	72.07	6.64	21
	Instructor	77.24	6.86	18
Grade	No feedback	67.85	6.64	18
	Computer	75.50	6.71	21
	Instructor	72.07	6.65	19

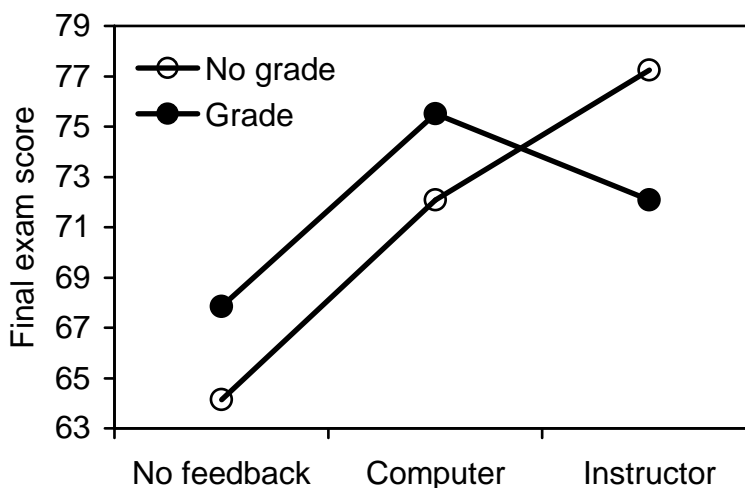*Note*. Adjusted means after controlling for the first exam score.

Figure 6. Mean final exam score as function of grade and feedback source for low ability

students.

No significant effects were found for grade, F (1, 103) = .275, p = .60; praise, F (1, 103) = .515, p = .48, or for interactions between grade and praise, F (2, 103) = 2.38, p = .13; praise and source of feedback, F (2, 103) = .24, p = .79, and among praise, grade, and source of feedback, F (2, 103) = .091, p = .91.

Table 13

Means and Standard Deviations of the Final Exam Scores by Source of Feedback, Ability Level, and Grade

		Low Ability			Medium Ability			High Ability		
		No Grade	Grade	Total	No Grade	Grade	Total	No Grade	Grade	Total
No	M	63.11	67.94	65.46	74.24	74.23	74.23	84.74	84.27	84.49
Feedback	SD	5.14	5.98	6.01	4.68	4.98	4.79	4.47	5.30	4.88
	N	19	18	37	42	35	77	19	22	41
Computer	M	72.05	75.71	73.88	80.78	78.16	79.54	90.00	87.32	88.49
	SD	7.81	9.09	8.58	5.57	4.66	5.29	4.87	5.14	5.14
	N	21	21	42	41	37	78	17	22	39
Instructor	M	78.06	72.32	75.11	82.00	78.33	80.23	89.79	87.81	88.76
	SD	7.52	5.10	6.94	6.04	6.18	6.33	4.65	3.89	4.35
	N	18	19	37	32	30	62	24	26	50
Total	M	70.98	72.19		78.73	76.86		88.25	86.54	
	SD	9.14	7.63		6.37	5.54		5.18	4.95	
	N	58	58		115	102		60	70	

Medium ability students

For the medium ability students, a significant effect for the source of feedback, F (2, 204) = 34.87, $p < .001$, η^2 = .26, was found. Pairwise comparisons revealed that students in the control condition scored significantly lower than those in either instructor ($p < .001$) or computer condition ($p < .001$) (see Table 13 for means). Additionally, significant differences were found between medium ability participants in the grade and no grade conditions, F (1, 204) = 7.9, $p < .001$, η^2 = .09. Students who were shown their

first exam grade scored lower than those who were not shown their grade. Means are presented in Table 13.

No significant effects were found for praise, $F (1, 204) = .063$, $p = .80$, or for interactions between grade and praise, $F (2, 204) = 2.67$, $p = .10$; grade and the source of feedback, $F (2, 204) = 2.34$, $p = .10$; praise and source of feedback, $F (2, 204) = .14$, $p = .87$, and among praise, grade, and source of feedback, $F (2, 204) = 1.96$, $p = .14$.

High ability students

For the high ability group, ANCOVA revealed significant effect for the source of feedback, $F (2, 117) = 18.13$, $p < .001$, $\eta^2 = .24$, with students in the control condition scoring significantly lower than those in either the instructor or computer conditions. No differences were found between the computer and instructor conditions, $p > .05$. Additionally, significant differences were found between the grade and no grade conditions, $F (1, 117) = 3.72$, $p < .05$, $\eta^2 = .05$. High ability students in the grade condition scored significantly lower than those in the no grade condition (see Table 13 for means).

No significant effects were found for praise, $F (1, 117) = .075$, $p = .78$, or for interactions between grade and praise, $F (2, 117) = .343$, $p = .56$; grade and feedback source, $F (2, 117) = .17$, $p = .85$; praise and the source of feedback, $F (2, 117) = 1.02$, $p = .36$, and among praise, grade, and the source of feedback, $F (2, 117) = .705$, $p = .50$.

Overall, the analyses showed that low ability students respond favorably to detailed feedback and are able to improve. However, when presented with a grade from the instructor, low ability students do not do as well as when they are oblivious to their first exam grade. At the same time, it appears that low ability students can handle a low

grade well if they believe it came from the computer. Both medium and high ability students were shown to respond well to detailed feedback coming from either computer or the instructor. Their performance, however, depended on whether a grade was presented, with those who received a grade scoring lower than those who did not.

Analyses of Differences in Motivation, Self-efficacy, and Affect

The third research question asked whether differential feedback affects students' personal characteristics of motivation, self-efficacy, and negative and positive affect. To answer this question, two 3 x 2 x 2 multivariate analysis of variances (MANOVA) were employed. The first MANOVA included self-efficacy and motivation as dependent variables, and grade, praise, and the source of feedback as independent variables. The second MANOVA was run with positive affect and negative affect scale scores as dependent variables, and grade, praise, and the source of feedback as independent variables.

For self-efficacy and motivation, multivariate tests were significant for the grade factor (the F statistic for Wilks' Lambda was F (2, 449) = 5.42, $p < .01$), and for the praise factor (the F statistic for Wilks' Lambda was F (2, 449) = 4.02, $p < .01$). To test the difference for both of the dependent variables, univariate analyses were performed for motivation and self-efficacy.

For motivation, the univariate results indicate significant differences in motivation levels between students who were praised on their performance and those who were not, F (1, 450) = 7.58, $p < .01$, $\eta^2 = .04$. Interestingly, students in the praise condition reported lower motivation (M = 47.29, SD = 7.66) than students in the no praise condition (M =

49.06, SD = 5.71). No differences were found between the grade and no grade conditions, F (1, 450) = .95, p = .33.

For self-efficacy, the results indicated a significant grade effect, F (1, 450) = 10.80, p < .01, η^2 = .08, with students who received a grade for the first exam exhibiting lower self-efficacy levels (M = 43.38, SD = 7.03) than those who were unaware of their first exam score (M = 45.47, SD = 6.36). No significant effects in the reported levels of self-efficacy were discovered for praise, F (1, 450) = .14, p = .70.

For positive and negative affect, multivariate tests were only significant for the grade factor, the F statistic for Wilks' Lambda was F (2, 450) = 7.03, p = .01. To test the difference for both of the dependent variables, univariate analyses were performed for both positive and negative affect variables.

Similarly to self-efficacy, there was a significant difference in negative affect depending on the presence or absence of grade, F (1, 450) = 14.09, p < .01, η^2 = .08. Students who received a grade for the first exam reported higher levels of negative affect (M = 25.27, SD = 7.68) as compared to those who did not receive their first exam grade (M = 22.72, SD = 7.12).

For positive affect, there were no significant effects for any of the independent variables or their interactions.

Overall, presence of grade was shown to have a significant effect on students' reported self-efficacy and negative affect. Students who received a grade had higher negative affect and lower reported levels of self-efficacy than their counterparts for whom their grade was unknown. Praise affected motivation, but in an unusual fashion,

with students presented with a laudatory statement reporting lower levels of motivation than those who were not.

Analyses of Differences in Perceived Helpfulness and Accuracy of Feedback

To answer the research question about differences in perceived helpfulness of feedback and perceived accuracy of feedback, a 3 x 2 x 2 MANOVA was employed. Perceived helpfulness and accuracy of feedback were used as dependent variables, and grade, praise, and the source of feedback as independent variables. Multivariate analyses revealed significant effects for the feedback source and for the interaction between praise and feedback source, the F statistic for Wilks' Lambda was $F(4, 900) = 87.10$, $p < .001$, and $F(4, 900) = 6.44$, $p < .001$, respectively.

Subsequent univariate analyses with the perceived accuracy of feedback as dependent variable showed a significant interaction between praise and feedback source, $F(2, 451) = 4.31$, $p < .05$, $\eta^2 = .04$. As shown in *Figure 7*, under the no feedback condition, students' ratings of the accuracy of feedback were higher for those who received praise (M = 3.70, SD = 1.50, for the praise condition, M = 2.91, SD = 1.50, for the no praise condition). The overall ratings of the feedback accuracy were relatively low for this group.

Under the instructor condition, students' ratings of the accuracy of feedback were relatively high, and were almost identical for the no praise and praise conditions, M = 5.94, SD = 1.50, and M = 5.96, SD = 1.50, respectively. Under the computer condition students' ratings of feedback accuracy were slightly lower when no praise was presented, M = 5.25, SD = 1.50, as compared to the ratings of students who received a laudatory

statement, M = 5.41, SD = 1.50. Means and standard deviations are presented in Table

14.

Table 14

Means and Standard Deviation of the Perceived Accuracy of Feedback by Praise and

Source of Feedback

		M	SD	N
No praise	No feedback	2.91	1.50	78
	Computer	5.41	1.50	79
	Instructor	5.94	1.50	76
Praise	No feedback	3.70	1.50	77
	Computer	5.25	1.50	80
	Instructor	5.96	1.50	73

Note. Adjusted means after controlling for the first exam score.
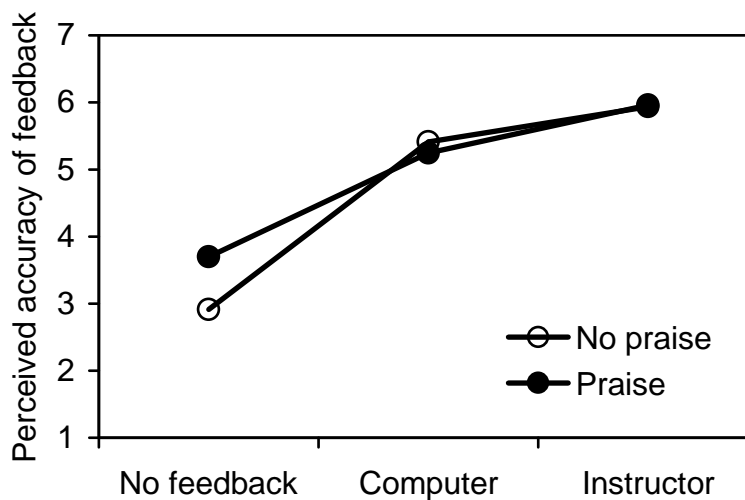

Figure 7. Mean perceived accuracy of feedback as function of praise and feedback

source.

Additionally, there was a significant effect for the source of feedback, F (2, 451) = 130.98, $p < .001$, $\eta^2 = .37$. A post hoc Scheffé analysis yielded a significant difference in accuracy ratings between instructor and computer conditions, $p < .01$, between instructor and no feedback conditions, $p < .01$, and between the computer and no feedback conditions, $p < .01$. Students who received their feedback from the instructor rated feedback as being more accurate (M = 5.95, SD = 1.07) than those who received feedback from computer (M = 5.33, SD = 1.42) or those who did not receive detailed feedback (M = 3.30, SD = 1.91).

Univariate analysis with perceived helpfulness of feedback revealed a significant effect for the source of feedback, F (2, 451) = 206.12, $p < .001$, $\eta^2 = .48$. A post hoc Scheffé analysis indicated a significant difference in helpfulness of feedback ratings between the instructor and computer conditions, $p < .01$, between the instructor and no feedback conditions, $p < .01$, and between the computer and no feedback conditions, $p < .01$. Students who received feedback from the instructor rated it as being more helpful (M = 6.06, SD = 1.07) than those who believed that feedback was computer-generated (M = 5.44, SD = 1.56) or those who did not receive detailed feedback (M = 2.79, SD = 1.76). Means and standard deviations are presented in Table 15.

Table 15

Means and Standard Deviation of the Perceived Helpfulness of Feedback by Source of Feedback

	M	SD	N
No feedback	2.79	1.76	155
Computer	5.44	1.56	159
Instructor	6.06	1.07	149

No significant effects were found for the interaction between grade and praise, F (1, 451) = .00, p = .98.

Overall, students rated feedback from the instructor as more helpful and accurate. Not surprisingly, students who received no detailed feedback reported the lowest levels of feedback helpfulness and accuracy. It is interesting to note that students in the no feedback condition, who were praised on their performance, had higher ratings of accuracy than those who were not praised on their performance.

Goal Orientation and its Effects on Students Performance

In order to test the hypothesis of whether students' goal orientation affects their responses to feedback, the following analyses were conducted. Firstly, students' responses to the 10 mastery and the 10 performance statements were summed individually to form a total mastery and a total performance goal score. From these scores, median splits were used to categorize participants into goal orientation groups. The maximum score on both the mastery and performance goal orientation scales was 70. Students who scored at or above 58 on the mastery scale were classified as having high mastery goals, and those who scored at or below 57 were classified as having low mastery goals. Students who scored at or above 51 on the performance scale were classified as having high performance goals whereas students who scored below 50 were classified as having low-performance goals. This procedure resulted in approximately 29% (n = 135) of the students being classified as having a high mastery/high performance goal orientation (multiple goal orientation), 26% (n = 120) as having a high mastery/low performance orientation (mastery orientation), 22% (n = 103) as having a low

mastery/high performance approach (performance orientation), and 22% (n = 104) as

having a low mastery/low performance goal orientation.

Table 16

Means and Standard Deviations of the Final Exam Scores by Source of Feedback for

Students with Different Goal Orientation

		Multiple Goal Orientation	Mastery Goal Orientation	Performance Goal Orientation	Low Mastery and Performance Goal Orientation
No Feedback	M	74.78	73.54	77.97	73.49
	SD	8.36	9.05	8.43	7.67
	N	51	37	32	35
Computer	M	79.69	80.67	80.76	79.63
	SD	8.05	9.09	7.24	8.06
	N	51	42	37	27
Instructor	M	81.70	82.27	79.91	82.98
	SD	8.03	8.19	7.66	7.85
	N	33	41	33	42
Total	M	78.33	79.02	79.61	78.91
	SD	8.60	9.48	7.77	8.79
	N	135	120	102	104

Secondly, the dataset was split on the goal orientation variable, and a 3 x 2 x 2

ANCOVA with the source of feedback (x 3), grade (x 2), and praise (x 2) as factors, and

the first exam grade as a covariate examined differences in the final exam scores for

students in each goal orientation group. Whenever ANCOVA revealed significant

differences among the three groups of the feedback source, pairwise comparisons were

performed between each pair.

Multiple Goal Orientation

For students with a multiple goal orientation, the analysis revealed a significant

effect for the source of feedback, $F (2, 122) = 18.69$, $p < .001$, $\eta^2 = .23$, with students in

the control condition scoring significantly lower than those in either the instructor (p <

.001) or the computer condition (p < .001) (see Table 16 for means and standard

deviations). Pairwise comparisons revealed no differences among students in the

computer and instructor conditions (p > .05).

There was also a significant praise by feedback source interaction, F (2, 122) =

3.22, $p < .05$, $\eta^2 = .08$; see *Figure 8*. In the no feedback condition, scores were higher for

those students who received praise, M = 75.36, SD = 5.20, than for those who did not

receive praise, M = 73.93, SD = 5.28. As shown in *Figure 8*, the overall scores were

relatively low for the no feedback group. Under the instructor condition, students' final

exam scores were relatively high, but they were lower when praise was added, M =

80.06, SD = 5.20, than when there was no praise, M = 81.08, SD = 5.21. Under the

computer condition students' scores were lower for those students who received praise,

M = 82.21, SD = 5.51, than for those who did not receive praise, M = 78.28, SD = 5.31.

Means and standard deviations are presented in Table 17.

No significant effects were found for grade, F (1, 122) = 2.65, $p = .11$, praise, F

(1, 122) = 1.57, $p = .21$, and for interactions between grade and praise, F (1, 122) = .01, p

= .91, grade and the source of feedback, F (2, 122) = .35, $p = .71$, and among grade,

praise and the source of feedback, F (2, 122) = 1.54, $p = .22$.

Table 17

Estimated Marginal Means and Standard Deviation of the Final Exam Grade by Praise

and Source of Feedback for Students with a Multiple Goal Orientation

		M	SD	N
No praise	No feedback	73.93	5.28	24
	Computer	82.21	5.51	24
	Instructor	81.08	5.21	19
Praise	No feedback	75.36	5.20	27
	Computer	78.28	5.31	27
	Instructor	80.06	5.20	14

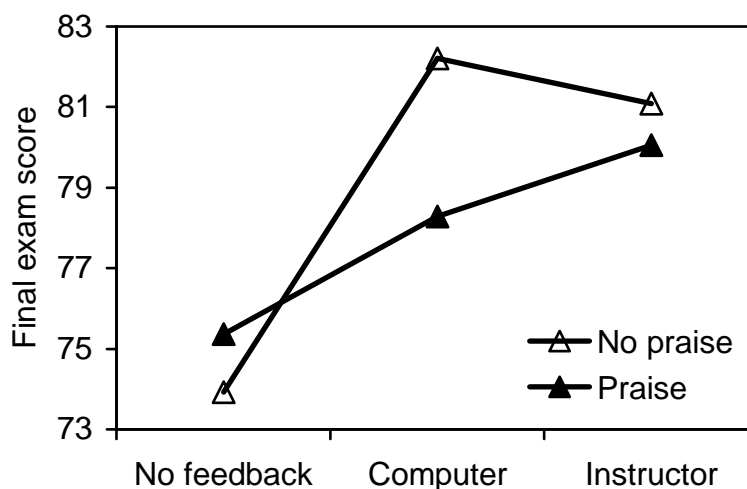*Note*. Adjusted means after controlling for the first exam score.

Figure 8. Mean final exam score as a function of praise and feedback source for students with a multiple goal orientation.

Mastery Goal Orientation

ANCOVA analysis showed a significant effect for the source of feedback, $F_{(2, 107)} = 13.64$, $p < .001$, $\eta^2 = .20$ (see Table 16 for means and standard deviations).

Students in the no feedback condition had a significantly lower average exam score than those in either the instructor (p < .001) or the computer condition (p < .001). Pairwise comparisons showed no differences among students in the computer and instructor conditions (p > .05).

Additionally, there was a significant grade by feedback source interaction, F (2, 107) = 5.89, $p < .01$, $\eta^2 = .12$; see *Figure 9*. In the no feedback condition, scores were higher for students' who received their grade, M = 76.39, SD = 1.24, than for those who did not receive a grade, M = 73.50, SD = 1.35. *Figure 9* demonstrates that the overall scores were relatively low for the no feedback group. Under the instructor condition, students' final exam scores were lower in the no grade condition, M = 73.50, SD = 5.58, than in the grade condition, M = 76.39, SD = 5.54. Under the computer condition, students' scores were higher in the no grade condition, M = 80.65, SD = 5.45, than in the grade condition, M = 79.28, SD = 5.49. Means and standard deviations are presented in Table 18.

Table 18

Estimated Marginal Means and Standard Deviation of the Final Exam Grade by Grade and Source of Feedback for Students with a Mastery Orientation

		M	SD	N
No grade	No feedback	73.50	5.58	17
	Computer	79.28	5.49	21
	Instructor	83.87	5.62	24
Grade	No feedback	76.39	5.54	20
	Computer	80.65	5.45	21
	Instructor	78.68	5.44	17

Note. Adjusted means after controlling for the first exam score.
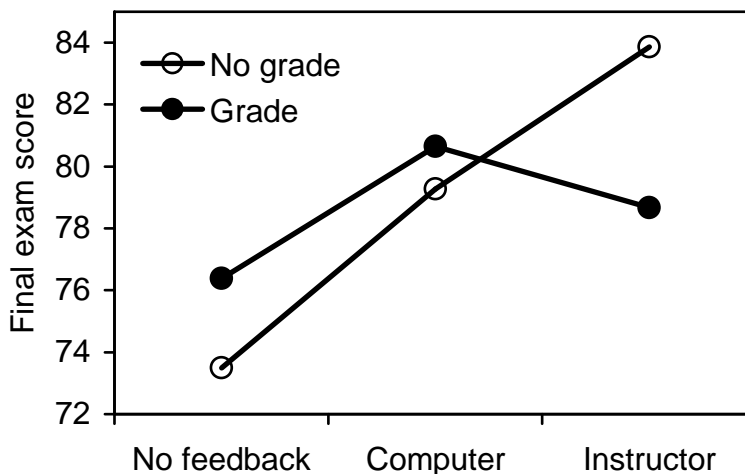
Figure 9. Mean final exam score as a function of grade and feedback source for students with a mastery goal orientation.

No significant effects were found for grade, F (1, 107) = .09, p = .76, praise, F (1, 107) = 2.32, p = .13, and for interactions between grade and praise, F (1, 107) = 1.08, p = .30, praise and the source of feedback, F (2, 122) = .37, p = .69, and among grade, praise and the source of feedback, F (2, 107) = 1.92, p = .15.

Performance Goal Orientation

For students with a performance goal orientation, the analysis revealed a significant effect for the source of feedback, F (2, 89) = 11.37, p < .001, η^2 = .20 (see Table 16 for means and standard deviations). Students in the control condition scored significantly lower than those in either the instructor (p < .001) or the computer condition (p < .001). No differences were found between the computer and instructor conditions (p > .05)

Additionally, the analysis revealed a significant interaction among grade, praise, and feedback source, $F (2, 89) = 4.02$, $p < .05$, $\eta^2 = .08$ (see *Figure 10*).

No praise condition
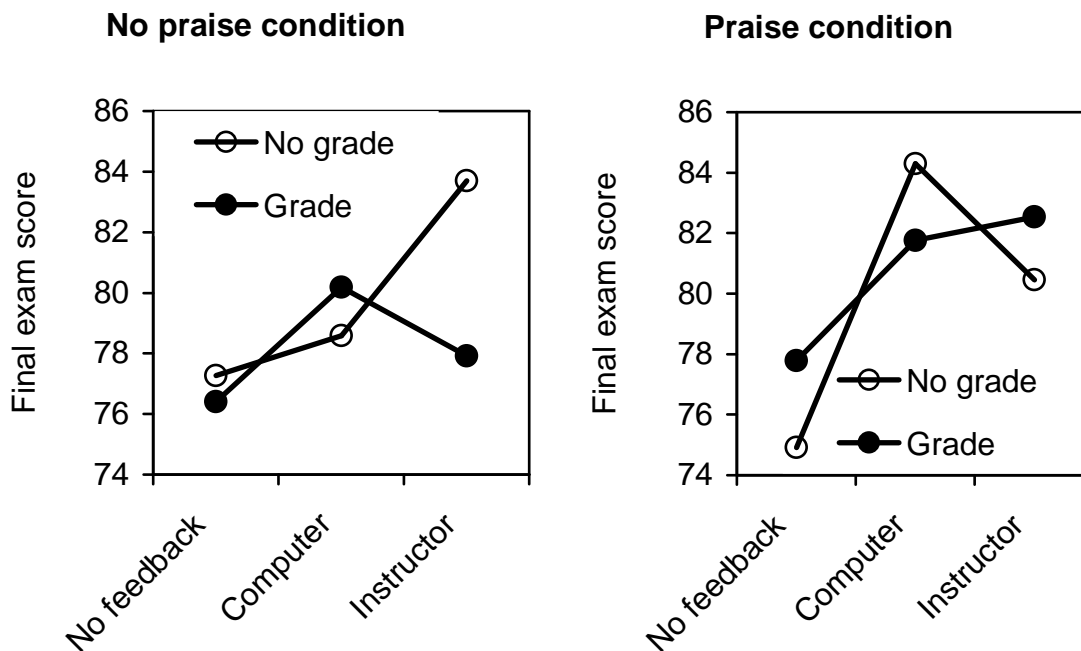
Praise condition

Figure 10. Mean final exam score as a function of grade and feedback source for students with a performance goal orientation; praise and no praise condition.

The three way interaction appears to a degree to reflect earlier findings. That is, in the no praise condition, there was a strong difference between students who received a grade perceived to come from the instructor (M = 77.92, SD = 4.39), and those who did not receive a grade (M = 83.69, SD = 4.38). The students who did not receive a grade performed better than those who did. The presence of praise with the grade ameliorated the negative effect of the grade to a degree. The computer feedback also produced interesting results, although somewhat difficult to interpret. The students receiving praise,

but no grade, showed the best scores (M = 84.30, SD = 4.38), followed by praise and a

grade (M = 81.76, SD = 4.35), then a grade with no praise (M = 80.18; SD = 4.36), and,

lastly, the no praise, no grade group (M = 78.58; SD = 4.35). In the no feedback

condition, students' scores were similar regardless of the praise and grade combination.

See Table 19 for means and standard deviations.

Table 19

Estimated Marginal Means and Standard Deviation of the Final Exam Score by Grade,

Praise and Source of Feedback for Students with Performance Goal Orientation

			M	SD	N
No grade	No praise	No feedback	77.27	4.36	8
		Computer	78.58	4.35	9
		Instructor	83.69	4.38	5
	Praise	No feedback	74.91	4.36	10
		Computer	84.30	4.38	9
		Instructor	80.45	4.35	5
Grade	No praise	No feedback	76.40	4.37	7
		Computer	80.18	4.36	9
		Instructor	77.92	4.39	11
	Praise	No feedback	77.79	4.49	7
		Computer	81.76	4.35	10
		Instructor	82.54	4.35	12

Note. Adjusted means after controlling for the first exam score.

No significant effects were found for grade, $F (1, 89) = .24$, $p = .63$, praise, $F (1,

89) = 2.04$, $p = .16$, and for interactions between grade and praise, $F (1, 89) = 1.92$, $p =

.17$, praise and the source of feedback, $F (2, 89) = 1.98$, $p = .15$, and grade and the source

of feedback, $F (2, 89) = .79$, $p = .46$.

Low Mastery/Low Performance

For the low mastery/low performance orientation group ANCOVA revealed a significant effect for the source of feedback, F (2, 91) = 22.15, p < .001, η^2 = .33 (see Table 16 for means and standard deviations). Students in the no feedback condition (M = 73.49, SD = 7.67) scored significantly lower than those in either the instructor (M = 82.98, SD = 7.85), p < .001, or the computer condition (M = 79.63, SD = 8.06), p < .001. No differences were revealed between the computer and instructor conditions (p > .05).

No significant effects were found for grade, *F* (1, 91) = 1.43, *p* = .24, praise, *F* (1, 91) = .07, *p* = .80, and for interactions between grade and praise, *F* (1, 91) = 2.66, *p* = .11, praise and the source of feedback, *F* (2, 91) = .38, *p* = .68, grade and the source of feedback, *F* (2, 91) = 1.23, *p* = .30, and among grade, praise, and the source of feedback, *F* (2, 91) = 2.77, *p* = .07.

Overall, the type of goal orientation was shown to lead to variation in students' responses to feedback messages. Students with all four types of goal orientation reacted positively to individualized comments, with students in the no feedback condition scoring lower than students in either the instructor or computer conditions. For students with low mastery and low performance goal orientation, the aforementioned effect was the only significant effect revealed. Mastery-oriented students reacted negatively to the grade if it was presented along with comments. For performance-oriented students, the most important finding was that a grade did not improve their performance unless paired with praise. Finally, students with a multiple goal orientation may have perceived praise as an unnecessary noise, taking their attention away from the task, and therefore scored higher in conditions in which a laudatory statement was omitted.

Analysis of the Focus Group Discussions

Analysis of the experimental data revealed intriguing patterns in students' responses to differential feedback. To further explore students' reactions, focus group discussions, held shortly after students completed their essay exams, were analyzed with the aim to substantiate the quantitative findings with students' accounts of their experiences. Examination of students' perceptions of various feedback messages provides a more comprehensive picture of their responses to instructional feedback and of its effects on their motivation and emotion. Such analysis helps to identify the kinds of feedback that best support learning.

Instructor feedback with a grade

This group consisted of eight students, all of whom received a grade and personalized comments with the understanding that the feedback was provided by the instructor. In addition to a grade and comments, four of the students received a general statement praising them on their performance.

"I loved the feedback," "the comments were really helpful," "I was like, wow, he's actually giving us feedback. Like, he read all of the essays! He's giving us feedback and it was actually cool" were among the first responses of the focus group participants to the most general question asked: "How did you react to the feedback?" When prompted, the participants expanded and talked about the importance and usefulness of comments for making revisions and ensuring improvement. One of the participants noted "I was relieved when I went through the comments. It felt nice to know exactly what I needed to do." This statement resonates with the views of the participants in this group.

Although there was a consensus about the effectiveness of personalized comments, the participants' reactions to the grade they had received were not uniform for this group. Some students described their main reaction towards their grade as "panic," "shame," "disappointment," and "anger." Students elaborated by saying that receiving a grade which was much lower than they had anticipated was discouraging, and it took time and special effort to convince themselves to move forward with their revisions. The students noted that the presence of comments made it much easier to resume their work on their essays. They knew exactly what to do, and tried to take each suggestion the professor provided and incorporate it into their essay. Interestingly, several participants emphasized that their negative emotion was directed at the instructor. A participant noted: "I got kind of mad at [the instructor]. I thought he was way too hard on me." Two other students concurred that receiving a grade made them think that the instructor was a tough grader, a sentiment leading to escalated feeling of helplessness and worry about their future performance.

Similarly, several participants reported feeling dissatisfied with their performance and feeling embarrassed in front of the instructor: "I thought, gosh, he [the instructor] thinks I am so dumb. I really felt terrible. I felt like I let him down." A student reported contemplating a change of major, reasoning that a poor performance on a writing task is indicative of her overall poor skills and inevitable failure in the writing domain. Apparently, feelings of incompetence were a common reaction after receiving a low grade. It appears the majority of emotional responses described by students could be labeled as negative. Students felt angry at the instructor, ashamed for letting him down, or simply incompetent. However, the availability of detailed comments nearly eclipsed

the initial strong reaction, as the participants reported working hard on improving their essays.

In those cases when the grade presented to the students was high, the participants expressed a different view on the matter. A student stated: "I got an 85 and I was satisfied so I just did some spelling corrections and that was it." Two other students agreed that getting a high grade served as an indicator of how much work needed to be done. "I knew I could make my work better, but why waste my effort if I already like my grade? So, I just made a few changes here and there and figured that it was already an improvement from before." Obviously, students were relaxed when they received a high grade, but in most cases chose not to invest a lot of effort into their seemingly satisfactory essays.

In terms of students' reactions to praise, those participants who were presented with a laudatory statement reported an overall positive attitude toward it. One of the high-scoring students indicated that receiving praise made her feel happy and enthused: "[The professor] addressed me by my first name, so I thought like he really knew me and really thought I did great. I wanted to do even better." Apparently, praise made students feel that the professor believed they could succeed. Praise also served as a buffer for students' self-efficacy in those cases when a poor grade was received. "I liked the praise comment. I first got mad at myself and him [the professor] for the grade, but then I thought he [the professor] gave me 68 just to push me more. I wasn't too devastated because he said I could do it."

Additionally, when discussing students' perceptions on the ideal feedback, the participants unanimously agreed that presentation of comments was the most important component. Detailed information on the mistakes and ways to improve them was deemed

as desirable and necessary for students' improvement. In regard to the grades, students indicated that low grades were generally discouraging, but served as useful indicators of how much work needed to be done in order to achieve a satisfactory result. In fact, the entire group contended that a grade in combination with comments and praise was the kind of feedback they would want to receive. They reasoned that individualized comments "tell you what to do," grades "tell you how much you need to do," and praise "makes you feel happy." Rephrasing the question about the ideal feedback to "What kind of feedback is ideal for your *learning*?" led to different responses. Some students noted that grades did not contribute in any way to their improvement, and therefore should be omitted. They saw a potential negative effect of both low and high grades, with the former leading to negative emotional reactions, and the latter reducing the effort they are willing to expend. Praise, on the other hand, was deemed beneficial to students' mood and attitude toward the task at hand. The participants agreed, however, that it was not crucial for their improvement.

Instructor Feedback without a Grade

There were nine participants in the second focus group. These students were presented with detailed descriptive feedback which they believed was provided by the course instructor. In addition to personalized comments, four of the participants were praised for their performance. Grades were not presented to anyone in this group.

This group had the highest degree of agreement in terms of their reported reactions to the feedback they received as compared to other focus groups conducted in the study. The participants were enthusiastic about the detailed comments they received,

and were appreciative of the clear guidelines on how to approach their revisions. A student noted: It was like going through a checklist. Fix this and this and you'll be fine. I knew my essay was good, and by fixing what I was told to fix I will only make it better. It was the best exam ever!" Other participants agreed, stating that the comments prompted them to stay focused on the task and helped them to feel productive at all times during the revision process. They "did not have to wonder what they did wrong," as the comments provided specific information on what needed to be modified in order to improve the essay.

Some students were initially alarmed by the abundance of comments, but they reported shifting into a more constructive mode soon after looking through the instructor's suggestions, as the following comment illustrates: "I had a lot of comments, so I thought, wow, I messed it up. But then I thought that whatever wasn't underlined must have been good, so as long as I do what you tell me to do, I'll do well on this exam." Other students reported feeling worried when they saw their essays with numerous corrections, but it was a transient feeling which was immediately substituted by positive emotional reactions: "I kind of, I was really confident at first about my essay, but then once I saw it, my morale dropped because I'm like, wow, this guy really chopped up my essay! But as I went along and as I finished it I was, like, now it's really-really good."

Students were pleased with the professor's commitment to their learning and were grateful for having instructions on how to proceed with their revisions. The following quotes illustrate this point: "I was amazed that he [the professor] gave me so much information… I thought, he's the best;" "I was, like, cool, he took time to help me better

my essay!" and "I never got so much feedback, it's so useful but nobody ever does it. I couldn't believe that he cared so much to do it for us." Obviously, the students responded positively to the feedback they received. They perceived the comments as the evidence of the instructor's commitment to their progress and as clear directions to what they needed to accomplish to improve their performance on the exam. These two themes emerged in the utterances of every participant of this group.

When asked about their reactions to praise, the four participants who had been presented with a laudatory statement reported feeling encouraged by it: "I thought, cool, he likes what I did and he thinks I can do better!" "Oh, I did a great job! I'm on the right track! So I didn't totally fail this." Students felt that praise was a professor's way to let them know that he believed they could do well on the exam. The participants noted that praise made them feel happy and removed any residue of worry which they may have had prior to receiving their feedback:

> I was worried that I failed the exam, but the praise made me feel like I can do this and I just have to fix some things. And then I went back and fixed the things that the professor noted. When I was leaving it made me feel better like it wasn't that bad I might do ok on this essay.

Interestingly, one student remarked: "I thought, maybe I already did great — after all he [the professor] said it — and now he just wants to push me. Maybe I didn't need to do all of it [the revisions] to get a good grade." Apparently, in this particular case praise led the student to conclude that his work may have been already good enough to receive a great score. Speculatively, praise may have depressed student's motivation to invest a lot of effort into his work on the essay.

The focus group discussed the idea of the ideal feedback. The students agreed that detailed comments were crucial for their improvement: "Tell me, like, specifically what

you want, because if you tell me what you want I can give it to you." The pivotal role of comments was clear to all. Additionally, the participants noted that praise may enhance the beneficial effect of detailed feedback. Students reasoned that praise would make them feel good about themselves and therefore, would lead to even better improvement. However, they concluded by saying that praise was not nearly as important as comments: "Praise without comments is not worth much." In regards to grades, students acknowledged the grades' potential to hurt their performance and motivation. They suggested that high grades would inevitably lead to reduced effort, and low grades would be very discouraging. At the same time, they noted that there could be a potential benefit in grades. Grades may inform students of how much work needs to be done. However, if the goal is to advance understanding (in this case, to learn to write the best possible essay), grades should not be provided.

Computer Feedback with a Grade

The third group was comprised of eight students who received a grade along with detailed feedback. Students were led to believe that both the grade and comments were computer-generated. In addition to the grade and comments, four participants received a general statement praising them on their performance and encouraging them to work to improve their essay.

In regard to specific feedback, students agreed that it was very helpful. The participants noted that when they first learned that their work had been evaluated by specialized software, rather than the instructor, they were cautious about the quality of the comments. However, their opinion changed as they began incorporating the

suggestions into their work. A student mentioned: "I was a bit iffy at first. But then I saw that the comments were so great and to the point. I didn't know that computers could be this sophisticated." Other students echoed by saying how amazed they were by the level of detail provided by the machine and the relevance of feedback to their essay.

Several participants mentioned that they were relieved when they realized that the computer graded their work. A student who received a low score remarked "I thought, thank God it wasn't the professor who read it [the essay] — it was so bad! I would have been mortified if he was the one who graded it." Apparently, students perceived the computer as being non-judgmental and impersonal which helped them focus on their work without worrying about their tainted reputation in the eyes of the course instructor. Similarly, some students felt that computer could have been more fair than the instructor when assigning grades and providing comments:

> I like the computer better because the idea of a computer is, I know there wouldn't be any bias. You can perceive things based on someone's name or how they use their words, and with a computer I feel like I'm pretty sure it would be programmed to not be biased.

Despite the undisputed quality of the comments, seven of the eight participants felt that some suggestions provided by the computer did not apply to their work. A student remarked: "I thought the computer didn't understand what I was trying to say. It told me to change things but it didn't make any sense. It would require major rewriting. So I still kept the same idea." Another group participant continued: "Some suggestions were weird so I figured I'd ignore them. Machines are not humans, so it [the computer] can't gauge my thinking the way a human would." Apparently, when in doubt, students chose to ignore the computer's comments, justifying their decision with potential flaws in the software.

Students' reactions to the grade they received were consistent with the aforementioned pattern. Some participants were skeptical of the fairness of their grade. Students reported thinking that their grade did not correspond to reality and was too low. They reasoned that the software may be inappropriate for assessing the complex task of writing, so it cannot gauge more advanced structures and their underlying meaning. "I didn't take it personally. It [the grade] was too low to be true. I didn't think it was my grade at all. I just kept revising and using my own reasoning to do it." Three of the participants, however, reported having different thoughts after receiving their grade. These students felt that the grade was fair because the machine was impartial when conducting assessment and generating the score. Obviously, students' views on the computer-provided grades and comments differed. Some students trusted the grade and the comments, reasoning that the machine was unbiased, whereas other students were very skeptical about the relevance of the feedback and the fairness of a grade, rationalizing that computers are not suitable for evaluating complex tasks. Skepticism of the quality of computer-generated comments in fact was apparent to varying degrees in the remarks of most participants.

Students who received praise did not report thinking seriously about it. The participants judged the encouraging comment to be "too generic," "dry," and "meaningless." They felt that the same comment was provided to every student in class, and therefore, carried no special meaning. With further probing, students noted that they would have liked to receive a more personal comment, as illustrated by the following remark: "It should include a name or something. Computers can insert names automatically. It would have made it seem more personal. Everyone likes being praised."

The participants agreed that although praise may not help with their performance, it does not hurt either. Moreover, personalized laudatory comments have a potential to increase morale and, therefore, should be presented along with the detailed descriptive feedback.

The kind of feedback that was identified by the participants as the most effective for learning was "specific comments with a grade and praise." The functions of each were identified as follows: Detailed feedback would provide information about specific mistakes and ways to correct them. Grades would tell students how much work needs to be done. Praise was deemed as having no particular purpose, but was judged as having a potential to elicit positive emotion. Some participant noted that they would only want to receive a grade if it is very high. In that case, they would not be discouraged, but at the same time would not be as motivated to exert additional effort and work on improving their work. Finally, most students stated that they would prefer receiving feedback from the instructor, as opposed to the computer. The instructor's feedback would naturally be more personal, accurate, and trustworthy, and therefore would be most effective in ensuring progress of learning.

Computer Feedback without a Grade

The fourth focus group was comprised of eight students who received detailed comments with the understanding that the feedback was computer-generated. In addition to comments, four participants had been praised on their performance. Grades were not presented to anyone in this group.

Group participants reported feeling pleased with the comments they received. Students agreed that clear guidelines provided by the computer helped them during the

review process. Not only were the comments instrumental in helping students make corrections, they also helped students concentrate on the task at hand, making them feel more confident that they could succeed. Students mentioned feeling relieved that there was a chance to rework their essay to ultimately get a higher grade. "I thought it was great. It gave me a chance to review. I basically looked them [the comments] over and knew I needed to revise my paper more and get a good grade. The comments were so to-the-point!" In regards to the credibility of the computer comments, many students remarked having original doubts but later finding the feedback to be useful and relevant. However, students admitted encountering comments that they felt did not apply, as the following statement illustrates: "The comments were great. It was so interesting that the machine is so smart. It was very useful, but some comments didn't really work, so I figured, oh, well, it's a machine after all. I should decide what to use." Other participants confirmed feeling that in some cases the computer was not quite "understanding" what they were trying to convey. Hence, they ignored some comments but incorporated those that they felt were relevant.

The participants who were praised on their performance were in accord in terms of their reactions to the praise comment. Students perceived praise as a machine-generated generic statement, as shown by the following remark: "I thought it was a comment that everyone got. The computer isn't warm and fuzzy, so I didn't think my essay was special. I just thought it was a general statement." Those students who were not presented with praise said that they would not have wanted to receive it. The participants felt that praise has a minimal, if any, value in their achievement. At the same time, students unanimously agreed that if presented by a person, praise may be useful for

enhancing positive affect and increasing motivation. Still, they noted that personalized comments were far more important for ensuring success, with praise serving as a "feel-good factor."

One of the students stated categorically, with universal head-nodding in agreement, that grades are unnecessary if a chance to review work is offered. Students explained that a grade is perceived as a "final step" and therefore is not conducive to motivating additional work on the assignment. "If I got a grade, I would be, like, OK, that's what I got. If you gave me an 85, I would look at the comments, but with my busy life, I would probably be, alright, I am satisfied with the grade, why stress about it?" Other student reaffirmed and expanded: "If it [the computer] gave me a 65, I would have panicked, but I probably would have put a lot in there. But if you want to give me suggestions, don't make me freak out with the low grade. I don't see the point." Obviously, students felt that personalized comments were sufficient and necessary for improvement. Grades, however, were perceived as having a potential to deplete effort and elevate anxiety, and were judged to be undesirable.

Students' views on the ideal feedback were similar to those expressed by participants of other focus groups. Students emphasized the pivotal role that detailed comments play for their improvement. They stressed that detailed feedback helps their learning, and having a chance to review and rework their assignments is beneficial for their progress. Computer comments were deemed to be very useful. However, students believed that the instructor's comments would be more trusted and, as a result, would lead to higher improvement. In regard to grades, students pronounced in unison that the only value of a grade is to reduce the amount of work in case of a high score. Even if

presented with numerous comments, a high grade would inevitably lead to decreased effort. Praise was said to be valuable for elevating mood, but not particularly effective for enhancing performance. The following comment summarizes the general attitude toward the ideal feedback:

> A chance to revise is so fair. I think I would like to get comments on every paper I write. Even when I saw that my paper was obliterated with marks on it and stuff, I still thought it was great. I could improve it. A grade would have freaked me out because I saw that much info… Praise is nice, but I don't care. I just want you to tell me what I need to do. This is a learning experience, after all.

No Detailed Feedback, Grade

The eight students who comprised this group had not been given any specific feedback on their work. All of the participants were presented with a grade, and four students were also provided with a laudatory statement.

The students' reactions to the feedback were not nearly as optimistic as those of the participants in the previous groups. "It was really bad," "I panicked," "I was shocked" were among the common remarks that participants shared. Students who received unsatisfactory grades without any guidance on how to improve their work reported feeing frustrated, as the following comment illustrates: "I worked really hard, and I got a 75. I was completely freaked out. You know it's low, you want to do something better but you didn't know what, so you didn't. I just moved some stuff around and left." Students felt helpless and craved any form of specific instructions. The majority of students reported not being able to considerably revise their essays. Rather, they claimed to have made minor adjustments to their work, due to a lack of information on their errors and ways to improve them. "I just corrected spelling. I didn't know what else I could possibly say," noted one of the participants.

Two of the participants received a high grade on their essay. They had a different reaction to the feedback as compared to their peers who got a lower grade. One student explained that "it was great to get an 85. I was satisfied so I resubmitted my essay without doing pretty much anything." The other high-scorer echoed, "I was glad to have been shown my grade. I was like, great, I did well, I don't need to revise." The students admitted, however, that if their grade had not been high enough to be considered satisfactory they would have been discouraged by it.

Four of the group participants received praise in addition to a grade, and their reactions to it varied. Two participants who had been praised on their performance did not find the comment to be useful. A student thought that praise was "a way of sugar coating what's coming your way. I got a 77 and then I got good praise, yeah, I didn't like that, it was weird." Another student agreed, noting, "I had a 73 and I didn't think he [the professor] was serious. I was just surprised to hear him say "you had a good start." It didn't really seem to reflect reality." Other participants had a different view on the matter: "I felt like, oh, I did a great job! I'm on the right track! So I didn't totally fail this." Among those students who did not receive praise, the opinions had a similar split. Some students wished they had gotten an encouragement, reasoning that it would have made them feel more confident about their performance, whereas other students said that praise would not have been instrumental. A student mentioned, "I didn't even want positive comments so much, just some sort of pointing out where the weak points are would be really helpful." This comment reflects this group's view on the feedback they consider being the most effective.

The participants agreed that having detailed comments would be of a great assistance, as "telling me exactly what to do is the best thing a professor can do." Students repeated that a grade on its own is not helpful. However, some of the participants suggested that a combination of a grade with the comments would constitute the ideal feedback:

> I think if it was comments and a grade it would have been the best. If I got a low grade, I would have appreciated it kind of, because I would be, all right, this is the grade I got, I need to improve it, definitely focus on the things they said I should improve upon.

However, the student mentioned that if he had gotten a higher grade, he "would be more hesitant to change anything," afraid to "make it worse." With the further exploration, the group derived certain contingencies associated with the presentation of a grade: "I think it would just depend, like, if someone got an 80 they'd probably know all the stuff they just don't have it all together. And then if they got that grade they can say "Ok, I just got to go back and put it all together." If someone gets a 60, that just means they didn't know what they were talking about, so the grade would make them get mad or panicked. If someone gets a 90, they're going to say "I'm not going to change anything." If it was me I wouldn't." Thus, students felt that in most situations a grade will tend to have a harmful effect on performance. The participants further elaborated by saying that for those courses that they were interested in and wanted to gain competence, grades would be discouraging. Conversely, for those courses that they dislike, a grade would be desired because it "just saves your effort." A student proposed that "if it's a passing grade, I would just resubmit it [the essay]. I wouldn't care even if I could improve." Students explained that "not knowing a grade in this case would be annoying, because I would be afraid that I failed, and would be forced to revise."

Students' comments indicate that they consider the usefulness of grades from two perspectives. On the one hand, if their goal is to get through the class with the minimal effort, a grade is judged to be helpful for telling them how much work they need to do. On the other hand, if the goal is to enhance understanding and improve learning, grades either make them nervous or prevent from investing a lot of effort. Thus, even if students are motivated to learn, they feel that grades may add an unnecessary constraint which may take their mind away from learning, and focus on energy and effort conservation: "Even if you want to learn more, you'd be like, why do it if you already have a 90." One of the participants voiced an opinion supported by all of the others, saying that when there is a choice between receiving no feedback at all, and receiving a grade, the latter alternative is definitely more appealing.

In regard to praise, students felt that praise would be a pleasant addition to the comments, but not a very useful form of feedback when presented alone. There was general agreement that more personalized and work-specific praise may be beneficial for students' motivation and a general sense of well-being. It was also noted that praise may soften the negative effect of a low grade by canceling out or mitigating the negative emotions which usually follow an unsatisfactory performance. However, when compared to the importance of comments, its value was deemed to be quite limited.

No Detailed Feedback, No Grade

The final group was conducted with eight students, four of whom received praise in the course of the experiment, and four who received no feedback of any kind. Interestingly, students' reaction to the absence of feedback was not overwhelmingly

negative. One of the focus group participants reported feeling grateful to receive a chance to revise his essay, as illustrated by the following remark: "I wish I could have gotten some feedback. Still, the second time I regrouped and remembered everything a little better. Revising really helps." Other students agreed, but added that it was frustrating not to have any guidance on what to change. They reported working on "whatever came to mind," "correcting stylistic and spelling errors," but not working so much on the content of their essay. Students felt that changing the content may lower their final score, so they chose not to make considerable revisions.

One of the participants interpreted the lack of any feedback as a clear indicator of the high quality of her work. The student commented: "I thought mine was really good. I had almost everything, like all the information that he [the professor] gave us in the lectures. I only deleted one sentence and added another, and left." Another student reported having the same reaction, reasoning that if grave mistakes were committed, the professor would have commented on them. Thus, he did not spend a lot of time on his revisions and left feeling satisfied with his performance.

Quite surprisingly, those students who were praised on their performance had stronger negative reactions than those who did not receive anything. A student noted: "It was a joke, like, "you made a great start, let's try to make it better," and nothing else! I didn't know how to make it better!" Another student echoed: "I thought it was just totally generic. It didn't help. It made me get more nervous, because I had no idea what to change." Obviously, the lack of clear instructions on how to improve work was very frustrating to these students. They knew they were expected to revise their work but were not provided any guidance on how to proceed with this task. Despite their frustration,

students who received laudatory comments reported that they spent more time on revisions than their counterparts who received no feedback. Students who were presented with praise received an explicit encouragement to continue improving their essay, whereas those who had no feedback were left to make this decision for themselves.

The participants of this group unanimously concluded that the ideal feedback should include specific comments. They felt that the chance to revise was useful, but that they lacked one important component — feedback. A participant commented: "If I'm going to get a chance to revise and get it back, I like to have some sort of idea of what I did wrong and what I should do to fix it." Students stated that if the detailed comments were impossible to be compiled, any form of feedback would be conducive to improvement. "Any information is better than no information," one of the participants shrewdly noted.

In regard to a grade, students felt that neither numeric scores nor letter grades were effective in promoting improvement, reasoning that "it makes you too anxious to actually know your grade." The participants agreed that "it's better just to critique your ideas, like, this is what you need to fix. If they give you a low grade it'll make you nervous. If you get a high grade, it's like I don't want to fix it." Similarly to the previous groups, the participants of this group were clear about their view on the potential negative effect of grades.

Additionally, students were in agreement that personalized comments represented the ideal form of feedback and were believed to lead to the best improvement. They also hypothesized that when praise was added to detailed feedback it enhanced the beneficial effect of the feedback. The participants speculated that praise would induce positive

affect, and comments would guide them through the revision process, thus resulting in the optimal progress. However, as one of the participants noted, "I don't care about praise. I want directions. Tell me what to do, and I will." This statement reflects the significance students assigned to constructive feedback as opposed to evaluative forms of feedback.

General Themes that Emerged in the Focus Group Discussions

The analysis of the focus group discussions allowed for the construction of a general picture of the personal experiences of the participants who had received different forms of feedback in the course of the experiment. Not surprisingly, students' responses differed both within each group and among the eight groups. The questions of the focus group discussions mirrored the main research questions of the study, with the addition of a question that explored students' views on the ideal feedback.

The participants who received detailed feedback from both the instructor and the computer emphasized the usefulness of comments for their performance. Students agreed that information about errors they committed and suggestions on how to correct them was necessary for their improvement. The reaction to the descriptive comments was predominantly positive for both computer and instructor groups. In some cases, students felt initially overwhelmed by the number of revisions they needed to make. They reported, however, that this feeling was transient and that they soon took a more productive stance of "knowing exactly what to do." For the instructor group, students felt that in addition to assisting them in the revision process, feedback was an indicator of the instructor's commitment to their progress. As a result, they were eager to make the

adjustments and improve their work. Computer-generated feedback was appreciated by the focus group participants and was generally deemed to be relevant and helpful. Interestingly, however, students in the computer group unanimously agreed that some of the comments they received did not apply to their work or were too outlandish to consider. Those in the instructor condition received essentially the same comments, but did not feel this way. This indicates that students judged the quality of the feedback based at least partially on its source.

Students who were not presented with detailed feedback reported feeling discouraged when they received a grade that they were not satisfied with, whether or not they received praise as well. They noted that having no guidance as to how to approach the task of making revisions was frustrating and often led to elevated negative affect. Conversely, students who received only a high grade were content and did not spend time trying to improve their work. Some of the students who did not receive any feedback at all interpreted the lack of comments as an indication of the high quality of their work. As a result, they chose not to revise their essays. Praise statements, albeit lacking clear directions on how to improve, were deemed to be more useful, for students who received an encouragement knew they were expected to revise their essay. The general conclusion inferred by the students was that any feedback is better than no feedback.

Students' reactions to grades were different for students across the six focus groups. Participants who received a grade from the instructor reported feeling angry when their score was low, but said that they still worked hard to incorporate the instructor's comments, thus improving their essays. Students whose grade was high enough to match their own standards admitted spending very little time considering the

instructor's comments. In the computer condition, low grades were received with skepticism by the majority of the participants. They were far less upset by their low score when it came from a computer, reasoning that the software was not capable of understanding the logic of their arguments, so the grade did not reflect reality. High grades, however, were thought to be fair so significant revisions were not made. The strongest negative reaction to the grade was reported by the students who did not receive anything in addition to the grade. In this case, they felt angry and helpless, not knowing what went wrong or how to improve it.

Praise elicited the most diverse responses from students. Under the instructor condition, the majority of students reacted positively to praise. They reported feeling happy and were encouraged to perform well on the exam. However, computer praise was dismissed by the majority of the focus group participants, though they said if it was made more personal, there could be a potential benefit to praise from a computer as a vehicle to increase their mood and motivation. When combined with the grade, praise appeared to have mitigated some of the negative effects of the grade. Students who received praise in addition to a grade felt less frustrated than did their counterparts who were presented with only a grade. The latter group expressed their desire to have been presented with a laudatory comment. General themes that emerged in the focus group discussions are listed in Table 20.

Students in all six focus groups concurred on their definitions of the ideal feedback. The focus group participants stressed the importance of detailed comments for their improvement and said that it was the most desired form of feedback they wished to receive. A grade was deemed to be unnecessary if the goal is to ensure progress in

learning. In every group, students admitted that receiving a high grade would inevitably lead to decreased effort, since there would be little room for improvement, as would receiving a low grade because it would be discouraging. Students noted, however, that for those courses that they strongly dislike, a grade would be a desired addition to comments. In such cases a grade would be a marker of how much work they need to do to obtain a satisfactory result. In regard to praise, students commented that it would be a pleasant addition to constructive feedback. It would elevate their morale and make them feel more confident. However, its role in students' performance was not considered to be critical.

Table 20

Prevalent themes of focus group discussions

Theme	Instructor Grade	Instructor No grade	Computer Grade	Computer No grade	Grade only	No feedback
Detailed feedback is useful, shows exactly what to do	√	√	√	√		
Some comments do not apply			√	√		
Low grades cause frustration	√				√	
High grades lead to reduced effort	√		√		√	
Personal praise is encouraging, elevates mood	√	√	√	√	√	
Praise softens the effect of grade	√				√	
Praise is useless			√	√		√

CHAPTER V

Discussion

This study attempted to shed light on the effects of differential feedback messages on students' performance, motivation, self-efficacy and affect. It also inquired into the potential differences in students' responses to feedback messages depending on their ability level and goal orientation. More specifically, it focused on determining the effects of grades, praise, and computer versus instructor provided feedback on students' performance and individual characteristics (motivation, self-efficacy, and affect), and tried to discern the main characteristic of feedback messages that produce optimal results. The experimental design of the study allowed for establishing direct influences among the variables and attempted to answer the question of whether a particular type of feedback leads to better improvement in students' scores. The authentic task employed in the study enhanced ecological validity and blocking based on students' first exam scores reduced sources of variability, thus leading to greater precision of the findings. The analysis of the focus group discussions allowed a more detailed picture of students' experiences to emerge.

The study contributes to clarifying a number of controversial areas in the field of assessment feedback. The most pervasive and strongest finding of the study was that descriptive feedback specific to individual work is critical to improvement. While less detailed forms of feedback, such as grades and general praise messages, lead to improved performance, it is to a significantly lesser extent. Moreover, when either grades or praise are added to descriptive feedback, they do not enhance the positive effect thereof, and do not contribute to improved performance. In fact, grades significantly reduce the

effectiveness of detailed comments, perhaps because of lowered motivation or negative affect. Even well-meant attempts to augment descriptive feedback with praise do not increase performance. It appears that the only benefit of praise reveals itself in those situations in which grades have been given. In this case, praise can soften the otherwise adverse effect of grades. Focus group discussions supported the finding of the experimental study. Students unanimously stressed that detailed comments were the most effective form of feedback. Grades were perceived as potential obstacles to student improvement, especially by those in the instructor condition. Praise was considered the least influential type of feedback, useful only to soften the demotivating effect of grades and to provide a general sense of well-being.

This chapter will discuss (a) differences in performance between students who received and did not receive detailed feedback, (b) differences in responses to computer and instructor provided feedback, (c) the effect of grades on students' performance on subsequent work, (d) the effect of praise on students' learning, and (e) differences in responses to feedback for students of different ability levels and goal orientations.

Descriptive Feedback and Its Effects on Learning

In formative assessment, educators attempt to move students from their current to their objective state of knowledge and skills, and employ methods that ensure the highest academic achievement (Black & William, 1998; Stiggins, 2005). Students in the study exhibited greatest improvement when they received feedback specific to their own work. This feedback pointed out students' mistakes, contained diagnostic annotations, and suggested possible strategies for improvement. The optimal feedback messages were

detailed, neutral in tone, and limited to direct statements about problems in students'

work and clear directions for solving them. Two experimental groups in the study

received descriptive comments and they significantly outperformed their counterparts in

the control group who were not presented with detailed comments. It did not matter to the

students whether their feedback was computer-generated or provided by the course

instructor — they effectively utilized directions for improving their work.

This finding is consistent with the body of literature on the subject. The meta-

analysis conducted by Klueger and DeNisi (1998) showed that correct solutions

feedback, as opposed to dichotomous judgments of correct or incorrect, led to greater

learning. Additionally, they found that neutral descriptive feedback, which conveys

information on how one performs the task and details ways to overcome difficulties, is

far more effective than evaluative feedback, which simply informs students about how

well they did and, consequently, carries a connotation of social comparison without

giving any guidelines on how to improve. Indeed, across the entire sample of the present

study, for students of all ability levels and different goal orientations, personalized

comments were imperative for greater improvement. The type of feedback, in this case,

detailed comments or lack thereof, accounted for 31 to 38 percent of variability in the

final exam scores. Thus, feedback appears to be a key to learning, and should be used to

promote students' success.

The importance of detailed feedback is especially pronounced for tasks that are

loosely-framed and do not have a clear right or wrong answer (Bangert-Drowns et al.,

1991; Roos & Hamilton, 2005). One of the main findings of the Bangert-Drowns' et al.

(1991) study was that feedback was less important for responses to simple or redundant

questions that can typify programmed instruction, but its effect was more prominent

when complex skills were tested. Providing elaborated information in the latter case led

to dramatic improvements, as opposed to simple correction of mistakes or providing any

form of evaluative feedback which did not appear to contribute to skill and knowledge

acquisition (Bangert-Drowns et al., 1991). No doubt, the essay writing task is not well-

defined. Not only did it require a strong command of the English language and good

writing skills, it also required deep understanding of numerous course-related concepts.

The complex nature of this task explains the crucial role that individualized comments

played in students' learning. Giving instructions on how to approach revisions and

providing information regarding committed errors and problems with students became a

fundamental requirement for students' improvement.

The success of detailed comments can also be explained through the lens of

information-processing theory, which emphasizes the importance of deep processing

when acquiring complex information (VanLehn, 1989). It seems that the detailed

comments provided in the study channeled students' attention towards relevant and

specific information, stimulated mental elaboration, and, consequently, boosted

performance. In addition, clear standards for self-evaluation created by individualized

comments may have provided information to students about the discrepancy between

their current performance and the standard they were trying to achieve, simultaneously

fostering their metacognitive skills.

In the focus group discussions, many students commented on the crucial

importance of the detailed feedback: "I thought it was really great because I knew exactly

what to do" or "I had a list of things I needed to fix, and I was calm because I knew that

if I do them all, my essay will become a top-notch work." At the same time, students who were not provided with descriptive feedback felt frustrated and helpless, and having received no guidance on how to make their essay better, did not put much effort into improving it. One student noted: "There was no way I could make it better — I had no clue what I did wrong" and "I just moved a couple of sentences around, but without knowing what exactly went wrong, I decided not to change much." Students' comments converge with the general findings of the study. The participants saw the benefit of feedback and realized its role in their achievement.

Overall, it seems apparent that providing specific personalized information about individuals' work and allowing them to make changes based upon this information leads to improvement of their performance. This study added to the corpus of literature and showed that the extent of students' improvement is to a large degree dependent on the presence or absence of individualized comments. In the educational system, knowing techniques that consistently work and lead to the best progress is a luxury. If detailed neutral comments represent one of the most powerful of these techniques, educators should put significant effort into providing them, and cultivate the resources to make it possible. Developing such comments may be an onerous task. However, if the goal of education is to advance students' progress, this practice will become worthwhile.

Differences in Responses Depending on the Perceived Source of Feedback

The main finding of the study that emphasized the beneficial effect of personalized feedback on students' performance can be further explored. The two experimental groups received feedback generated in the same fashion with the only

difference being that in one group students believed that their comments came from the computer, and in the other group students thought that their feedback came from the course instructor. This study revealed that students' improvement in performance was nearly equivalent for both feedback conditions. The presentation of meaningful comments, regardless of their source, was shown to help students learn. This finding appears to provide partial support for the "computers as social actors" (CASA) paradigm, suggesting that people may be unconsciously perceiving computers as "intentional social agents," and because of this, computer-provided feedback will tend to elicit the same or very similar responses from individuals (Nass et al., 1996; Nass et al., 1999).

The support the present study gives to the CASA paradigm is only partial, because although students' exam scores were quite similar for both computer and instructor conditions, differences in patterns of students' responses to feedback were consistently observed. Participants in the instructor condition, for instance, outperformed those in the computer condition when only comments were provided. However, when grade was presented along with comments, their scores were lower. The scores of their counterparts in the computer condition were the same regardless of whether their grade was presented or not. This interaction will be discussed in more detail in the following section.

The competing paradigm, which proposes that computers are generally perceived as neutral tools (Earley, 1988; Lepper et al., 1993), did not find support in either the experiment or in the focus group discussions. According to this perspective, computers tend to be viewed as neutral and unbiased sources of information. Thus feedback received from computers is more trusted by individuals. Quite contrary to this viewpoint, the

analysis of students' perceptions of accuracy and helpfulness of feedback revealed that

students rated the instructor's feedback as being more accurate and helpful than

computer-generated feedback. This finding converged with students' comments made

during the focus group discussions. A student mentioned: "The instructor knows us and

he is more personal than the computer is. The computer is helpful, but I was suspicious of

whether its comments make sense. It is not a person after all. It can't really think."

Similarly, a student remarked:

> I felt kind of like awed that a computer can do all that. I also felt like some of the
> feedback didn't apply. It was telling me to add this thing but I couldn't apply that
> to my life, to my subject, so I just ignored it and just kept going.

Considering that all feedback comments were verified for relevance by the

experimenter and the course instructor and mirrored the comments presented to students

in the instructor conditions, it seems apparent that students' in the present study felt

cautious about the accuracy of computer-provided comments. At the same time, it did not

prevent them from using the information to improve their work.

Computer assisted instruction, use of hypermedia, and sophisticated learning

environments have become an indelible part of modern instructional practices. Providing

feedback to their users is one of the main functions of these complex systems. The study

showed that, notwithstanding the higher perceived accuracy of instructor's feedback,

students' need for guidance and assistance may be addressed with equal success by both

computer and instructor-generated feedback. In both cases, a successful outcome is

contingent upon the relevance and meaningfulness of feedback. It is possible, however,

that in some situations skepticism of computer feedback may be quite strong, and

therefore, computer feedback may not be as effective as human provided comments.

Overall, it seems that as long as the feedback message encourages "mindfulness" in students' responses (Bangert-Drowns et al., 1991), students will treat computers as equals to humans and will use computer feedback to improve their work. This conclusion is consistent with the CASA perspective. However, skepticism articulated during the focus group discussions along with the different patterns of responses for computer and instructor conditions indicated that students do not treat human and machine generated feedback as the same. In fact, quite contrary to the "computers as neutral tools" perspective, students felt that in some cases, computer-provided suggestions for improvement were faulty and irrelevant.

The Effects of Grades on Students Learning

Letter grades and numeric scores are the most common types of feedback that students receive in a typical classroom (Marzano, 2000). With minimal time expense, grades provide a convenient summary of students' performance and inform all interested parties of students' achievement. Airasian (1994) identified five main functions of grades: (1) *administrative,* dealing with decisions concerning matriculation, retention and entrance into college; (2) *guidance,* by helping teachers, parents, and counselors provide direction to students; (3) *instructional planning*, by informing teachers about students' level of attainment in order to group them for instruction; (4) *feedback*, to provide students with information about their progress and achievement, and (5) *motivational*, to encourage students to try harder.

Within the realm of formative assessment, functions 4 and 5 of Airasian's (1994) list should be most pertinent to educators. Grades, therefore, should facilitate students'

learning by influencing their performance and motivation. The findings of the present study are indicative of the opposite trend. The evidence in this study was that students who received a grade scored significantly lower on the exam than their counterparts who were not shown their first exam grade. One of the explanations comes from the feedback intervention theory proposed by Klueger and DeNisi (1996). They suggested that optimal feedback should direct individuals' attention towards the task and towards the specific strategies that would lead to achievement of desired outcomes. Letter grades or numeric scores, being evaluative in nature and carrying a notion of social comparison, tend to turn students' attention away from the task and towards the self, thus leading to negative effects on performance (Kluger & DeNisi, 1996; Siero & Van Oudenhoven, 1995; Szalma et al., in press).

The attention diverted from the task to an individual's perceptions of self inevitably leads to reallocation of cognitive resources. Contemplating one's success or failure may subsequently impede effective performance due to competition for cognitive resources (Kanfer & Ackerman, 1989). Students' comments appear to agree with this explanation. One of the focus group participants shared: "I didn't know how I did till I got my grade. It just caused me to panic a lot. I couldn't focus on the essay, I thought of how badly I did." Another student echoed: "I was upset because I thought I did a lot better. I stared at it for, like, fifteen minutes before I could start making some changes. I kept thinking that I failed the exam." Speculatively, these comments indicate that presentation of a grade depleted students' cognitive resources, and therefore hindered their problem-solving behavior.

In a similar vein, attention to the self elicited by the presentation of a grade could activate affective reactions. Klueger, Lewinsohn, and Aiello (1994) argued that feedback received by individuals gets cognitively evaluated with respect to harm or benefit potential for the self and for the need to take an action. The appraisal of harm versus benefit is reflected in the primary dimension of mood (pleasantness), and the appraisal of the need for action is reflected in a secondary dimension of mood (arousal) (Kluger & DeNisi, 1996). The relationship between the two dimensions is not linear, as a potential threat to the self may instigate high activity on the student's behalf. At the same time, it may debilitate students so they cannot act.

The affective measure administered in this study tackled the arousal dimension of mood. High positive affect was indicative of high arousal, and high negative affect was indicative of depression and behavior inhibition (Crawford & Henry, 2004). The results indicated that students who were shown their grade scored significantly higher on the negative affect scale than their counterparts who did not receive their grade. The following comment captures the student's reaction to grade: "I saw my grade and froze. I can't really improve that much from 55 [referring to the score]. I am going to fail it. I felt quite mad." Possibly, frustration followed by helplessness prevented them from effectively carrying out their revisions and succeeding on the exam.

The negative effect of grades on students' performance can also be explained through their influences on students' self-efficacy. Generally, self-efficacy, or beliefs about one's competence, is known to be influenced by prior outcomes (Bandura & Locke, 2003). Feedback, therefore, has a potential of affecting self-efficacy. The present study revealed that presentation of grade resulted in decreased levels of self-efficacy.

Students who were not shown their grade reported higher levels of test-specific self-efficacy than those to whom a grade was provided. One student who received her grade remarked: "I felt, oh, my gosh, I am such an idiot. I am a terrible writer, and should change my major [from English] to something else."

In addition to being a consequence of performance, self-efficacy is consistently shown to be an antecedent thereof. Beliefs about one's competency were found to lead to greater use of diverse learning strategies, increased effort, sustained persistence, and higher attainment on a variety of tasks (Bandura, 1997; Lee & Klein, 2002; Linnenbrink & Pintrich, 2003; Schunk, 1990, 1995). Therefore, a grade that caused students' self-efficacy to plummet, could have subsequently negatively affected their performance on the second exam.

The adverse effect of grades can also be discussed using the following motivational dichotomies: task-involvement and ego-involvement orientation (Nicholls, 1984), and intrinsic and extrinsic motivation (Graham & Wiener, 1996). Task-involved individuals tend to be concerned with achieving mastery of a skill, and ego-involved individuals strive to score highly or attempt to avoid failure. In a similar way, intrinsically motivated students perform tasks for personal satisfaction, whereas extrinsically motivated students sustain their effort in order to achieve a reward. Using Nicholls' (1984) paradigm to guide her study, Butler (1988) demonstrated that grades induced an ego-involved motivational orientation, and undermined students' interest in the task, their enjoyment in performing it, and reduced the effort they were willing to expend to complete it. Butler and Nissan (1986) emphasized that grades decreased intrinsic motivation, depressed creativity, and fostered fear of failure. Although measures

of intrinsic and extrinsic motivation and task versus ego-involvement were not
administered in the present study, it is tempting to speculate that grades indeed led
participants' to concentrate on their final exam score thus reducing their interest in the
actual task, which led to a decreased desire or ability to try and improve their work.

Feedback is known to serve either an informative or a controlling function in
students' behavior (Bandura, 1982a; Lepper, 1981). In the former case, feedback
provides information about students' mastery, whereas in the latter case, it is used as a
reinforcer or punisher of students' behavior and represents an incentive to engage (or not)
in tasks (Stipek, 2002). Marzano (2000) states that the most important purpose of grades
is to provide information to students and, if referencing for grading is content-specific,
letter grades and numerical scores will lead to an increase in students' performance. He
postulates that if students have a clear understanding of the requirements of the task and
if grading is based only on students' achievement and effort, students can increase their
level of knowledge and understanding based on grades alone. Although plausible, this
view does not find support among researchers in the field, and neither did it find support
in the present study. Many researchers agree that grades are perceived by students as
controlling rather than informative (Elawar & Corno, 1985; Stipek, 2002). In fact, focus
group participants commented that grades were "useless if you want to improve,"
because they "don't tell me what I did wrong," and "don't give me any information on
how to fix my problems." Lack of specificity inherent to a grade does not allow for
effective remediation of one's work. As Roos and Hamilton (2005) note, information is
too deeply encoded in a grade for it to lead to appropriate action.

The overall negative effect of grades on students' performance is apparent. Nonetheless, subtleties in grades' effects on learning were found and need to be examined. Significant interactions were found between grades and feedback source and grades and praise. The first interaction revealed that presentation of a grade in the no feedback, instructor feedback, and computer feedback conditions led to different outcomes in students' performance. Because computer feedback is not a reasonable option in many courses, and because differences between grade and no grade conditions were less noteworthy for this group, the discussion will primarily focus on the differences between the no feedback and instructor feedback conditions.

In the no feedback condition, in which students did not receive descriptive feedback, presentation of a grade was actually beneficial to students' outcomes. Students who were presented with their grade exhibited somewhat higher performance than those who did not receive any kind of information. It should be noted, however, that students who did not receive detailed comments demonstrated significantly inferior performance than those who were presented with individualized comments.

These results converge to some extent with the findings of Butler and Nisan's (1986) study. In it, the researchers found that students who received grades performed better on a simple quantitative task than their counterparts who did not have any feedback on their performance. However, when a convergent thinking task was employed, students' performance was nearly identical for both groups. Although the essay writing task was very different from either of the tasks employed in the aforementioned study, it may be that students involved in the present study were more familiar with their task than students who had to carry out divergent thinking exercises. In the latter case, they may

have needed more support and guidance, and a simple score on their performance was absolutely meaningless to them. When writing an essay, however, a grade could have been construed by students as an indicator of *how much* work needs to be done. Although it did not provide any information of *what* exactly needs to be changed and how to go about it, the grade certainly served some informative function. It appears to be the case that any feedback is better than no feedback at all.

However, more feedback is not always better. In the instructor feedback group presentation of a grade resulted in significant differences between the two conditions. Students who received only detailed personalized comments from the instructor performed better on the exam than did their counterparts who received a grade in addition to the comments. The classic work of Page (1958), commonly cited when the effect of grades needs to be proclaimed as positive, or, at least neutral, demonstrated that optimal feedback included both comments and grades. Obviously, the results of the present study are indicative of the opposite contingency. The finding of this study bolsters the line of research carried out by Butler (1988), Butler and Nisan (1986), Elawar and Corno (1985). These studies demonstrated that feedback consisting of grades and comments led to significantly lower improvement than comments alone.

Grades appear to undermine students' interest relative to comments, to short-circuit students' thinking, and prevent them from using detailed feedback to improve their work. Both anecdotal evidence and research findings suggest that grades are perceived as potent sources of control over learning (Stipek, 2002). Students may perceive personalized comments as useful sources of information, but presentation of a grade proclaims the instructor's control over their learning, refocuses their attention on

the self and the quantitative aspect of learning, and reduces their desire or ability to

improve their learning (Butler & Nisan, 1986; Kohn, 1999).

There is a stunning irony in the following statement from one of the focus group

participants:

> I would definitely want both grades and comments. Comments tell you what to do, and grade tells you how much you want to do. I got 85 on my exam and I was fine with it. So, I just made a couple of corrections and left. I think grade is very useful. Why waste time doing corrections if you are already satisfied with your work?

It seems quite obvious that from this student's perspective, the only purpose of his

grade was to help him determine how much work he needed to do. If the grade was high,

he did not put the instructors' comments to good use and just left, perfectly content with

his original score. The goal of education — to help students' learning progress — was not

reached. In those cases when students received low grades and were informed of their

grades, their frustration, anxiety, and general negative affect caused by the grade

decreased their utilization of the detailed comments. Students felt that their course grade

was jeopardized (to a degree), and there was too much pressure for them to adequately

proceed with their corrections. To conclude the discussion of the potential detrimental

effects of grades on students' performance, the following quote effectively captures the

view on the matter that students hold. In it, a focus group discussant described the ideal

feedback that he would like to receive as well as possible effects of grades on his

performance.

> I think that you don't really need a grade, just so long as there's a critique in there like if there's a lot of comments or errors or anything, if they circled things and put actual punctuation in there, you kind of know you didn't do that good. And the grade, if it was in the 90's, I wouldn't have changed anything because I would have thought "I'll just make it worse." So I would have just resubmitted my work

so five minutes later you would've have seen me walking out. If it was very low, like in the 60's, I would have left because there's no hope for me anyway.

Although it is hard to disagree with the convenience and effectiveness of grades when used for summative purposes, the formative function of grades as tools that lead to progress in learning is quite dubious. In some educational settings, however, presenting a grade is a requirement. As a result, figuring out ways to do so with the least damage to students' achievement and, hopefully, with added benefit to their performance is crucial for educators across all academic environments. The possible solution to this quandary is presented below.

The Effects of Praise on Students' Learning

The present study attempted to clarify the effect of praise on students' performance, motivation, self-efficacy, and affect. Praise is a controversial topic, with some researchers arguing that praise promotes learning by raising positive affect and self-efficacy (Alber & Heward, 2000), while others stipulate that it leads to depletion of cognitive resources by taking attention away from the task and focusing it on aspects of the self (Baumeister et al., 1990; Kluger & DeNisi, 1996). This study did not reveal any differences in performance among students who did or did not receive praise on their performance. The general effect of praise on students' exam scores may be deemed as neutral. Comments and grades, alone and in combination, have a stronger influence on students' performance, with praise adding to and modifying their effects.

The only outcome measure directly affected by praise was motivation. The effect of praise was quite interesting, if not surprising. Students presented with praise reported lower levels of motivation as compared to their counterparts who were not praised on

their performance. Recall that student's motivation was measured after they had finished

their work, up to two hours since the time that they received their praise. Therefore, the

group differences found indicate that this type of feedback had a relatively stable effect

on the level of motivation. This is an intriguing finding as no studies known to date have

shown that praise negatively affects students' motivation. In fact, in most research

inquiries, praise is typically conceptualized as a condition that should enhance interest

and motivation by providing non-controlling, self-enhancing information about one's

competence (Deci et al., 1999; Delin & Baumeister, 1994; Ilies & Judge, 2005). The only

study that somewhat agrees with the finding here was conducted by Butler (1987). The

researcher demonstrated that students' receiving praise on their performance reported

high levels of ego-involvement, decreased levels of task-involvement, and higher

perceptions of success while exhibiting modest performance on a task as compared to

students who were not praised on their work. The motivation measure administered in the

present study did not tackle different types of motivation. It is possible that this general

motivation measure corresponded to the task-involvement measure employed by Butler

(1987), and therefore elicited similar responses. Students presented with praise were not

as interested in the task and were not as motivated to try harder, believing perhaps that

they had achieved enough. This supposition could be confirmed if students' performance

reflected it; however, praise appears to have a less direct, rather, a mitigating effect on

students' performance.

One of the findings of the present study was a significant interaction between

grade and praise. Students who received a grade without praise on their work scored

lower than those who were presented with an encouraging statement in addition to a

numeric score. Meanwhile, in the no grade condition, presentation of praise resulted in lower students' final exam scores as compared to those who received neither grade nor praise.

The preceding section presented the argument supporting the negative effect of grades and numerical scores on students' performance. According to the Feedback Intervention Theory and findings of Baumeister's et al. (1990) study, praise and grades should affect students' performance in a similar fashion. Both grades and praise can be expected to make individuals self-conscious thus leading to disruption of skilled performance. In other words, attention to the self resulting from praise and grades may rob cognitive resources that otherwise would be committed to the task. Only in a simple task, in which fewer resources are needed for its completion, would praise have a neutral or positive effect on performance (Baumeister et al., 1990).

The findings of the present study are not as straightforward. Adding praise to the picture was shown to soften the detrimental impact that grades have on students' learning. Grades and praise appear to carry the opposite signs in terms of their effects on performance, thus canceling each other out. Apparently, presentation of praise in addition to a grade mitigated the negative impact of the grade on self-efficacy. Students' perception of competence could have gone down with a grade, but praise elicited hope and belief in one's abilities, thus softening the effect of the former feedback regimen. Additionally, grades, normally perceived as strong sources of control over learning, usually result in negative outcomes, whereas the self-enhancing non-controlling nature of praise reduces the detrimental influence of a grade. Therefore, in situations in which grades must be presented to students, educators should consider accompanying it with

meaningful praise. However, it should be reiterated that when neither grades nor praise are presented, students' scores on the exam were the highest.

Hence, if educators have an option to choose, personalized comments without praise or grade appear to be presented as an optimal form of feedback leading to the highest increase in students' learning when a task is going to receive additional work by a student. Consider this illustrative comment that a focus group participant made: "Just comments, tell me what I did wrong, where I could change it. Just comments and error marks."

Difference in Responses to Feedback as Dependent on Students' Ability Level

Several researchers propose that students' responses to feedback messages may depend on their ability levels (Black & William, 1998). To date, very few studies have examined the differential effects of feedback on students' performance or motivation for students of different ability levels. Butler (1988) showed that presentation of a grade on its own or in combination with any other information leads to a significant decline of interest in performing the task for low achieving students. High achievers maintained their level of interest despite presentation of the grade. No other differences in responses were found between the low and high achieving groups.

In the present study, low, medium, and high ability students showed a significant increase in scores when presented with detailed comments. Once again, this finding attests to the fact that information regarding mistakes along with suggestions on how to improve them is a key to student achievement. It did not matter what their original grade was, students who were offered feedback specific to their own work found ways to

incorporate it into their essay and improve their results. Feedback accounted for 28% of variance in the final exam score for the low ability students, and for 26% and 24% for medium and high achieving students, respectively. Thus, the overwhelmingly positive effect of personalized comments permeated the entire sample, irrespective of students' ability levels.

Although detailed comments were conducive to learning in students' of all abilities, some differences in students' responses to feedback were found between the low ability group on the one side, and medium and high ability groups on the other. Students of high and medium ability performed differently when a grade was and was not presented. Under the grade condition both ability groups scored lower on their essay as compared to students who did not receive their grade. As suggested in preceding sections, a grade appears to undermine the effort that students' are willing to put forward in order to improve their work. Receiving a satisfactory grade may prevent students from channeling their effort towards further mastery of their work; rather, their focus on the quantitative aspect of learning leads them to give up before they can perfect their work. In focus group discussions students consistently noted that, after seeing a satisfactory grade, they "didn't have to do that much" and that grade showed them "how much work they needed to do to pass with a minimal effort." Thus, it is particularly important to note that presentation of a grade is especially damaging for higher ability students. There is always room for improvement even for the most successful students. Pushing them an extra step is likely to yield the best results when individualized, personally-tailored feedback is presented. A grade on its own or in combination with other forms of feedback prevents them from performing to the best of their ability.

Interestingly, however, no overall differences between the grade and no grade conditions were found for the low ability students. Instead, there was a strong grade by feedback source interaction which revealed that students who received neither specific feedback nor grade scored, on average, 15 points (out of 100) lower than those who received instructor's comments without a grade. The difference in performance between students who received their grade and those who received their grade along with the comments is far less dramatic and constituted an average of 4 points. Apparently, informative feedback is crucial for low scorers. Students who received a grade outperformed students who received no feedback and no grade. Thus, in this case the low grade led to positive effort on the part of the students. However, students who received detailed feedback and a grade performed less well than those who only received the detailed feedback. It seems to be the case that any kind of information is better than none, but that when detailed feedback was given, adding a grade hindered performance of the low performing students. Interestingly, the students in the computer feedback and grade condition outperformed those without the grade. It may be the case that the computer based grade was viewed as being less judgmental or personally directed than the instructor based grade. Students' comments reflect this speculation, with many focus groups participants mentioning that seeing their grade made them think that they could not possibly improve their work to earn a passing score. The grade caused them "to panic," "to freeze," or "to get angry at myself and the professor." Any of these emotions could hinder students' improvement on the task at hand, and obviously, they did.

Goal Orientation and Responses to Feedback

Goal orientation has been consistently found to have an effect on students'
learning, and it has been suggested that the type of goals that individuals have may lead
to variation in their responses to feedback (Black & William, 1998; Ilgen & Davis, 2000).
Mastery goals direct individuals to seek opportunities to reach new levels of competence
and master complex tasks (Dweck, 2000). Performance goals drive individuals to seek
and maintain a positive image of their ability (Pintrich & Schunk, 2002). Recent studies
have differentiated the construct of performance goals into approach and avoidance.
Individuals with a performance approach goal orientation are characterized as being
motivated to outperform others, whereas individuals with a performance avoidance goal
orientation strive to avoid failure or to appear incompetent in comparison to others
(Elliot, 2005; Pintrich & Schunk, 2002). The measure employed in the present study
only captured performance-approach goals; students' performance-avoidance goals were
not measured. Recently, multiple goal orientation was deemed to be common, with some
students being equally motivated by mastering a skill as by the chance to outperform
others (Elliot, 2005; Mattern, 2005). Informed by this proposition, this study investigated
responses to feedback of students with four different goal orientations: performance,
mastery, multiple goal orientation, and, finally, students' with an undetermined locus of
goal orientation.

For students with each of the four types of goal orientation presentation of
detailed comments played a crucial role in performance. To reiterate a unifying theme of
this discussion, individualized comments played the deciding role in students'
improvement. Students concerned with the grade, or mastery of a skill, or both at the

same time, scored significantly higher if they were shown what exactly they did wrong

and presented with suggestions on the steps they needed to take in order to improve. For

students with an undetermined locus of motivation, this was the only significant finding.

No other differences were revealed for this group of students.

Students with a mastery goal who were presented with both comments and a

grade performed significantly lower than those who were presented with comments

unaccompanied by a grade. Additionally, for mastery-oriented students who were not

presented with detailed comments, presentation of a grade led to increased performance.

The latter finding is easy to explain by taking into account characteristics of typical

mastery-oriented students. For these students, reaching new levels of competence is

imperative, and anything that aids them on this quest will generally be embraced (Ames,

1992). Grades, therefore, are perceived as information that informs them of their distance

from their desired level of mastery. This explanation does not support the first finding

though. The scores of students with a mastery goal orientation were lower when

comments were presented by the course instructor and contained a grade, as compared to

the exam scores of students' who were presented with detailed feedback without a grade.

According to Ames (1992), mastery goals should have reduced pressure associated with

trying to "look competent," and freed cognitive resources allocated to worrying, so

grades should not have caused any damage to students' performance. Obviously, the

findings of the present study do not agree with this picture. To venture a conjecture,

mastery-oriented students may indeed care about their competence on a task, and when

placed in a purely "learning" situation where no grades are assigned, perform quite well.

However, when they are placed in a situation in which grades are emphasized, mastery-

oriented students do not feel as driven to gain proficiency in a task at hand. In other words, it is quite possible that there were situational changes in goal orientation. Ward et al. (2004) discusses differences between dispositional and situational goal orientations, maintaining that goal orientation can be successfully manipulated by situational characteristics. The authors suggest that people set goals in a given situation, which results in a state, transient, situation-driven goal orientation. Thus, differential feedback messages could have inadvertently manipulated students' goal orientation, bringing students to care more about their grade and performance relative to others.

A complex interaction among grade, praise, and the source of information was found for students with a performance goal orientation. To avoid unnecessary complexity, the following discussion will focus on the patterns that seemed to be the most meaningful and relevant to educators. First and foremost, students with a performance goal orientation responded differently to a grade and praise presented by the instructor than they did to those presented by the computer. Students' scores were higher for those who received both a grade and praise and lower for those who received a grade without a laudatory statement. Surely, a grade coming from the professor may have made performance-oriented students self-conscious and could have increased their anxiety, resulting in their inability to use comments for further improvement (Ilgen & Davis, 2000). Praise received from the instructor may have served as a buffer that softened the impact of the grade on students' feelings of competency. As a result, students' perception that the instructor believed they could succeed on a task, resulted in a more significant improvement on the exam.

In the computer condition, however, grade did not affect students' performance when presented on its own. The computer may have been perceived as just a machine, which is not reliable and is not capable of grading students' work. Additionally, a grade coming from the machine may not have carried the otherwise strong notion of social comparison. The machine is not capable of relating a student's performance to other students; this information is purely mechanical and therefore less threatening to an individual's self. At the same time, the performance of students who received computer comments accompanied by praise and grade scored lower than those who received computer comments, praise, but no grade. Apparently, for the performance group, praise from the machine was not perceived as credible, and therefore did not alleviate the detrimental impact of grade on their sense competency.

However, despite the complexity of the interaction of the different feedback sources for the performance group orientation, the findings may be reduced to a quite uncomplicated statement that resonates with the general findings discussed in the present section. One way to understand the result for the performance group is to consider the following: Grades do not encourage improvement unless coupled with a message that shows that the teacher believes in the student's ability to improve.

Finally, for students with a multiple goal orientation, an interaction between praise and the source of feedback was found. Students who were not presented with descriptive feedback benefited from praise. However, when a laudatory statement was presented along with comments from either the computer or the instructor, students' scores were lower than for those students who received comments without being praised on their work. Multiple goal orientation students combine the need to score highly with

the need to gain competence. When no information is provided on students' performance or improvement thereof, students focus on praise, which makes them feel more competent and motivated to proceed with revisions. However, when comments are provided, they strive to perfect their work, with praise being perceived as an unnecessary distraction.

Overall, consistent with the literature on the subject, the type of goals that individuals hold was shown to lead to variation in their responses to feedback messages. Students with all four types of goal orientation reacted positively to individualized comments, but several differences in responses were revealed across the four groups. Contrary to the predictions that could be inferred from the existing literature, mastery-oriented students in some cases reacted negatively to a grade, but only if the grade was presented along with instructor comments. For performance-oriented students, the most important finding was that grades did not improve their performance unless paired with praise. Finally, students with a multiple goal orientation may have perceived praise as an unnecessary noise, taking their attention away from the task, and therefore scored higher in conditions in which a laudatory statement was omitted.

Limitations

Some potential limitations of the study should be noted. The study participants were college students attending classes at Rutgers University and The College of New Jersey enrolled in the general psychology course taught by the same instructor. The courses in both schools followed the same syllabus, used the same text, and had the same requirements. The fact that the same professor delivered lectures, responded to students'

queries, and was perceived by all students as the source of feedback, allowed for legitimate and meaningful comparisons across schools, and removed the unnecessary variation which would have resulted if two or more instructors taught the courses. However, the instructor's personality may have contributed to the results of the study. The course instructor was one of the highest rated professors in both schools, well-known for his striking ability to establish close rapport with students. The results of the study, therefore, may have been influenced by the fact that the students regarded their professor's opinions highly and sought to excel to obtain his approval. As a result, grade and praise may have had a stronger effect on students' performance. It should be noted, however, that the study demonstrated significant differences in students' responses across all conditions, and although a potential factor, the instructor's personality may not fully explain the revealed variation in students' learning progress.

It must be considered that the use of questionnaires in any study inevitably imposes certain limitations on the study. Students may not think deeply when they complete a questionnaire: they may misunderstand answers or provide socially desirable responses. In the present study students were asked to fill out several instruments immediately before they started working on their essay exam, or soon after they completed writing or revising their essays. Possible stress associated with the exam situation may have resulted in students' perfunctory answers to the questionnaire questions. It is possible that they rushed through questions in order to proceed with the exam. However, information obtained in the course of the focus group discussions that corroborated with the findings of the experiment may alleviate the concerns normally associated with questionnaire studies.

One of the feedback conditions in the study involved presentation of praise. The decision was made to use a standard laudatory comment differentiated according to three levels of the quality of students' work. No main effects were found for the praise factor. It is possible that none of the three levels of praise were strong enough to induce emotional responses that are commonly reported in the literature (Baumeister et al., 1990; Delin & Baumeister, 1994; Henderlong & Lepper, 2002). Laudatory comments that are more detailed and personal could have induced a broader range of responses from the participants. At the same time, interaction effects were found between praise and grade, as well as praise and feedback source, which indicate that the praise manipulation was successful.

The sample of the present study was comprised of college students who were relatively uniform in their age, with the majority of the participants being first-year students. Generalizing the results of the study to wider populations should be approached with caution. Conversely, the fact that the main experimental task was a part of a normal learning experience, and was approached by participants seriously as a regular course exam, contributed to the robustness of the findings.

Another potential limitation concerns the setting of the current study. Students worked on their exams in a computer lab. The lab was equipped with fifty computers placed relatively close to one another. As a result, in some cases students could glance into their classmates' screens and notice the kind of feedback their peers received. A few students were confused as to why their classmates were presented with detailed comments, while all that they were shown was a grade or an encouragement. In such cases, both the instructor and the experimenter carefully reiterated the nature of the

experiment, encouraging the participants to improve their work as much as they can based on the feedback they received. Such occurrences were not common. In most cases, however, students were seated according to their condition, and therefore had no access to other students' feedback.

Finally, the experimental task involved students working on an essay and then coming back a week later to revise their work based on the feedback provided at that time. In other words, the feedback was used to monitor and improve performance on an assignment carried out over a relatively brief period. The students were not assessed later, and they were not given a similar task at a later time. Therefore, the present study does not allow for inferences concerning the long-term effect of feedback on students' writing.

Directions for Future Research

Assessment is an ingrained part of any academic environment. Ideally, assessment of students' learning should contribute to their learning and academic attainment. Understanding the optimal ways to provide feedback to students and the most effective forms thereof is critical for success of any academic enterprise. The present study demonstrated that certain commonly used practices may hinder students' performance. In order to gain a more complete picture of the effects of differential feedback on students' performance, additional work can build upon these findings. First and foremost, the present study looked only at students' ability to communicate comprehension of an aspect of psychology through their writing. This is important as the ability to convey one's thoughts in writing is pertinent to every discipline. However, the mechanism through which feedback affects students' performance, as well as students' patterns of responses

to it may not be the same for all domains. Future studies should investigate the effects of feedback on students' performance on tasks coming from different domains of knowledge and academic domains.

Future research should explore reactions to feedback by students' of different ages and academic backgrounds. The sample of the present study consisted of college students, predominantly in their first year of college. It would be useful to investigate whether the results of the present study would hold with pupils in elementary, middle, and high school. It would also be interesting to find out whether presentation of praise is equally effective for students of all ages and whether patterns of responses are the same or similar across all ages and academic levels.

Research might also inquire into the long-term effect of feedback on students' learning progress. The essay writing task employed in the study involved students working on an essay and then coming back a week later to revise their work based on the provided feedback. It allowed for the isolation of the immediate effects of feedback on students' performance. Exploring the cumulative effect of feedback on students' learning by providing them with various kinds of feedback during the course of a semester or a year would aid in developing a more complete view of the potential effects of feedback on learning. Additionally, it would be useful to consider whether presentation of a certain type of feedback leads to differences in performance in those situations in which students do not receive a chance to revise and resubmit, but are asked to complete a similar assignment later.

Another area of investigation that may prove fruitful for future research concerns the role of individual characteristics in determining students' responses to feedback.

There is ample evidence suggesting that personal dispositions may affect students' reactions to feedback as well as that feedback may affect students' individual attributes. Studies have examined the role of feedback either influencing students' goal orientation (Latham & Locke, 1991), self-efficacy (Bandura & Locke, 2003), motivation (Kluger & DeNisi, 1996), affect (Ilies & Judge, 2005), or eliciting differential reactions from students depending on these individual characteristics. The findings of these studies, however, have been inconsistent and even contradictory (Kluger & DeNisi, 1996). In the course of the present research, measures of motivation, affect, goal-orientation, and self-efficacy were administered to the research participants. However, the motivation, self-efficacy, and affect measures were taken after feedback was provided. This allowed for examining how differential feedback affected these variables, but did not allow for the investigation of how these variables influenced how the feedback information was received. Future studies may attempt to examine whether efficacy levels lead individuals to exhibit differential responses to feedback messages, as well as investigate the interplay of goals and self-efficacy as factors influencing individuals' reactions to feedback.

It is also plausible that successful self-regulators may be more prone to effectively use information presented to them in a feedback message, and therefore improve upon it (Butler & Winne, 1995). Collecting self-report measures and latency indicators of self-regulation would help to answer questions on whether students' current level of self-regulation predicts their reaction to feedback messages, and whether a specific type of feedback leads to higher levels of self-regulating behavior.

Overall, the exact mechanisms through which feedback messages impact students performance and personal dispositions should be examined in future research inquiries.

Corroborating evidence from studies conducted across various domains of knowledge with students of different ages and levels of academic attainment would assist in understanding more fully the effect of feedback on learning and would allow researchers to make important additional conclusions about optimal feedback practices.

Summary

This study attempted to fill in the gap in the current understanding of differential effects of feedback on students' performance, motivation, affect, and self-efficacy. It also endeavored to uncover whether students of different ability levels and various goal orientations would respond differently to feedback messages. The authentic learning task contributed to the ecological validity of the study, and the classroom context insured that the participants approach the task with all due seriousness of a regular course exam. The current study is among the few that were conducted in an authentic learning environment. The findings, therefore, deserve careful attention from both researchers and practitioners.

In order to test the potential effects of feedback on students' performance, a valid assessment of their work was needed. The use of the ETS-provided assessment software along with the two highly-calibrated human raters ensured proper evaluation of students' work. Custom-made software was used to present feedback to students and allowed the control necessary to implement the design of the study. No studies known to date have used this level of complexity in both the design and the depth of assessment of students' products.

Additionally, a broad range of conditions allowed for isolating the effects of specific forms of feedback individually and in combination. Focus group discussions

aided in the interpretation of the experimental findings and showed students' perspectives regarding the preferred forms of feedback. The fact that for the most part the data obtained through both research paradigms converged serves as a form of triangulation and enhances the robustness and credibility of the results.

The most condensed conclusion of this inquiry is as follows: Detailed, specific, descriptive feedback which focuses students' attention on their work, rather then the self, is the most advantageous kind of information that should be provided to students. The benefit of such feedback occurs at all levels of performance. Evaluative feedback in the form of grades may be helpful if no other options are available, and can beneficially be accompanied by some form of encouragement. At the same time, grades were shown to decrease the effect of detailed feedback. It appears that this occurs because it reduces a sense of self-efficacy and elicits negative affect around the assessment task.

A quote of one of the focus group participants summarizes the discussion:

I wish I had all tests in high school and college like this [referring to the essay exam] because they didn't give you any feedback. I always hated grades because they don't tell you anything. I'm sitting there [at the exam] thinking this is great; I can fix anything I messed up on. I can make it [the essay] better… So I don't think grades or praise would be good. Just comments, tell me what I did wrong, where I could change it. Just comments and error marks.

This statement mirrors the findings of the study with uncanny precision.

REFERENCES

Airasian, P. W. (1994). *Classroom assessment* (2 ed.). New York, NY: McGraw-Hill.

Alber, S. R., & Heward, W. L. (1997). Recruit it or Lose it! Training students to recruit positive teacher attention. *Intervention in school and clinic, 32*, 275-282.

Alber, S. R., & Heward, W. L. (2000). Teaching students to recruit positive attention: A review and recommendations. *Journal of Behavioral Education, 10*, 177-204.

American Association Advancement of Science (AAAS). (1998). *Blueprints for Reform*. New York: Oxford University Press.

Ames, C. (1992). Classrooms: goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.

Anderman, E. M., & Johnston, J. (1998). Television News in the Classroom: What Are Adolescents Learning? *Journal of Adolescent Research, 13*, 73-100.

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology, 62*, 148-156.

Annett, J. (1969). Feedback and Human Behaviour. Baltimore, Maryland: Penguin.

Archer, J. (1994). Achievement goals as a measure of motivation in university students. *Contemporary Educational Psychology, 19*, 430–446.

Assessment Reform Group. (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: University of Cambridge, School of Education.

Attali, Y. (2004). *Exploring the feedback and revision features of the Criterion service*. Paper presented at the National Council on Measurement in Education Annual Meeting, San Diego, CA.

Bandura, A. (1982a). The Self and Mechanisms of Agency. In J. Suls (Ed.), *Psychological Perspectives on the Self* (pp. 3-39). Hillsdale, NJ: Erlbaum.

Bandura, A. (1982b). The Self and Mechanisms of Agency. In J. Suls (Ed.), *Psychological Perspectives on the Self,* (pp. 3-39). Hillsdale, NJ: Erlbaum.

Bandura, A. (1986). *Social Foundations of Thought and Action: a social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*, 87–99.

Bandura, A., & Wood, R. E. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision-making. *Journal of Personality and Social Psychology, 56*, 805-814.

Bangert-Drowns, R. L., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Baumeister, R. F., Hutton, D. G., & Cairns, K. J. (1990). Negative effects of praise on skilled performance. *Basic and Applied Social Psychology, 11*, 131-148.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-68.

Black, P., & William, D. (2003). In praiseof educational research. *British Educational Research Journal, 29*(5), 623-637.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (Eds.). (1971). *Handbook on the formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.

Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning. Finding a balance between learning goals and ego-protective goals. In B. J. Boekaerts, P. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 417-450). San Diego, CA: Academic Press.

Bouffard, T., Bouchard, M., Goulet, G., Denoncourt, I., & Couture, N. (2005). Influence of achievement goals and self-efficacy on students' self-regulationn and performance. *International Journal of Psychology, 40*(6), 373-384.

Boulet, M. M., Simard, G., & Demelo, D. (1990). Formative evaluation effects on learning music. *Journal of Educational Research, 84*, 119-125.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Criterion Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. *AI Magazine, 3*, 27-36.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245-281.

Butler, R. (1987). Task-Involving and Ego-Involving Properties of Evaluation: Effects of Different Feedback Conditions on Motivational Perceptions, Interest, and Performance. *Journal of Educational Psychology, 79*, 474-482.

Butler, R. (1988). Enhancing and undermining intrinsic motivation; the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*, 1-14.

Butler, R., & Nisan, M. (1986). Effects of No Feedback, Task-Related Comments, and Grades on Intrinsic Motivation and Performance. *Journal of Educational Psychology, 78*, 210-216.

Button, S. B., Mathieu, J. E., & Zajac, D. M. (1996). Goal orientation in organizational research: A conceptual and empirical foundation. *Organizational Behavior and Human Decision Processes, 67*(1), 26-48.

Cameron, J., & Pierce, D. P. (1994). Reinforcement, reward, and intrinsic motivation: a meta-analysis. *Review of Educational Research, 64*, 363-423.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review, 97*, 19-35.

Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology, 93*, 43-54.

Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback on enhancement of academic self-concept. *Journal of Educational Psychology, 83*, 17-27.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245–265.

Creswell, J. W. (2003). *Research design: Quantitative, qualitative, and mixed methods approaches.* (2nd ed.). Thousand Oaks, CA: Sage.

Crooks, T. J. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research, 58*(4), 438-481.

Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles, Policy and Practice, 9*(2), 237-252.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 627–668.

Delin, C. R., & Baumeister, R. F. (1994). Praise: More than just social reinforcement. *Journal for the Theory of Social Behaviour, 24*, 219–241.

Derryberry, D. (1991). The Immediate Effects of Positive and Negative Feedback Signals. *Journal of Personality and Social Psychology, 61*(2), 267-278.

Dev, P. C. (1997). Intrinsic motivation and academic achievement: what does their relationship imply for the classroom teacher. *Remedial and Special Education, 18*, 12-19.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048.

Dweck, C. S. (2000). *Self theories: Their role in motivation, personality, and development*. Philadelphia: Psychology Press.

Earley, P. C. (1988). Computer-generated performance feedback in the Subscription-processing industry. *Organizational Behavior and Human Decision Processes, 41*, 50-64.

Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: changing teacher behavior a little rather than a lot. *Journal of Educational Psychology, 77*, 162-173.

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52-72). New York: The Guilford Press.

Epsten, M. L., Epsten, B. B., & Brosvic, G. M. (2001). Immediate feedback during academic testing. *Psychological Reports, 88*, 889-894.

ETS. (2006). ETS Criterion: Topics. Retrieved October 5, 2006, from ETS Web site: http://www.ets.org/Media/Products/Criterion/topics/topics.htm.

Ferdig, R. E., & Mishra, P. (2004). Emotional responses to computers: Experiences in unfairness, anger and spite. *ournal of Educational Multimedia and Hypertext, 13*(2), 143-161.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: a meta-analysis. *Exceptional Children, 53*, 199-208.

Gagne, R. M. (1977). *The conditions of learning*. New York: Holt Rhinehart and Winston.

Gipps, C. (1994). *Beyond Testing. towards a theory of educational assessment*. London, UK: Falmer Press.

Graham, S., & Wiener, B. (1996). Theories and principles of motivation. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 63-84). New York: MacMillan.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition and Emotion, 4*, 269–288.

Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in Children's Learning: An Experimental and Individual Difference Investigation. *Journal of Personality and Social Psychology, 52*, 890-898.

Guskey, T., & Bailey, J. (2001). *Dveloping grading and reporting systems for student learning*. Thousand Oaks, CA: Crown.

Hanson, W. E., Creswell, J. W., Plano Clark, V. L., Petska, K. S., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology, 52*(2), 224-235.

Harackiewicz, J. M., Barron, K. E., Elliot, A. J., Carter, S. M., & Lehto, A. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest in making the grade. *Journal of Personality and Social Psychology, 73*, 1284-1295.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., & Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology, 92*, 316-330.

Henderlong, J., & Lepper, M. R. (2002). The Effects of Praise on Children's Intrinsic Motivation: A Review and Synthesis. *Psychological Bulletin, 128*(5), 774-795.

Ilgen, D. R., & Davis, C. A. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology: An International Review, 49*(3), 550-565.

Ilies, R., & Judge, T. A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology, 90*(3), 453-467.

Jolly, J. B., Dyck, M. J., Kramer, T. A., & Wherry, J. N. (1994). Integration of positive and negative affectivity and cognitive content specificity: Improved discrimination of anxious and depressive symptoms. *Journal of Abnormal Psychology, 103*, 544–552.

Jussim, L., Yen, H. J., & Aiello, J. R. (1995). Self-consistency, self-enhancement, and accuracy in reactions to feedback. *Journal of Experimental Social Psychology, 31*, 322-356.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integration/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*, 657-690.

Kanouse, D. E., Gumpert, P., & Canavan-Gumpert, D. (1981). The semantics of praise. In J. H. Harvey, W. Ickes & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 3, pp. 97–115). Hillsdale, NJ: Erlbaum.

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

Kluger, A. N., Lewinsohn, S., & Aiello, J. (1994). The influence of feedback on mood: Linear effects on pleasantness and curvilinear effects on arousal. *Organizational Behavior and Human Decision Processes, 60*, 276-299.

Kohn, A. (1993). *Punished by rewards: the trouble with gold stars, incentive plans, A's, praise, and other bribes*. Boston, MA: Houghton Mifflin.

Kohn, A. (1999). From grading to de-grading. *High School Magazine, 6*(5), 38-48.

Krapotkin, E. G., Korol, V. S., & Elevanova, I. A. (2002). *Accountability: Insuring Quality of Higher Education. Department of Educational Standards and Evaluation*. Retrieved. from.

Kulhavy, R. W., & Anderson, R. (1972). Delay-retention effect with multiple choice tests. *journal of Educational Psychology, 63*, 505-512.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in Written Instruction: The Place of Response Certitude. *Educational Psychology Review, 1*(4), 279-308.

Lajoie, S., & Derry, S. (1993). *Computers as cognitive tools*. Hillsdale: Lawrence Earlbaum Associates.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50*, 212-247.

Lee, S., & Klein, H. J. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology, 87*, 1175-1182.

Leontyev, A. N. (1981). *An outline of the evolution of the psyche: Problems of the Development of the Mind. [Эволюция психики: проблемы развития разума].* Moscow: Progress Publishers.

Lepper, M. R. (1981). Intrinsic and extrinsic motivation in children: Detrimental effects of superfluous social controls. In A. Collins (Ed.), *Aspects of the development of competence: The Minnesota symposia on child psychology* (Vol. 14, pp. 155-214). Hillsdale, NJ: Erlbaum.

Lepper, M. R., Henderlong, J., & Gingras, I. (1999). Understanding the effects of extrinsic rewards on intrinsic motivation-Uses and abuses of meta-analysis: Comment on Deci, Koestner, and Ryan (1999). *Psychological Bulletin, 125*, 669–676.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 75-106). Hillsdale: Lawrence Earlbaum Associates.

Lerner, R. M. (2002). *Concepts and Theories of Human Development* (3 ed.). Mahwah, NJ: Lawrence Erlbaum Publishers.

Linn, R. (2000). Assessments and Accountability. *Educational Researcher, 29*(2), 4-14.

Linn, R. L., & Miller, M. D. (2005). *Measurement and Assessment in Teaching* (9 ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading and writing quarterly, 19*, 119-137.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Sciences in the Public Interest, 14*(1), 47-53.

Marzano, R. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum and Development.

Mattern, R. A. (2005). College Students' Goal Orientations and Achievement. *International Journal of Teaching and Learning in Higher Education, 17*(1), 27-32.

Meyer, W.-U., Mittag, W., & Engler, U. (1986). Some effects of praise and blame on perceived ability and affect. *Social Cognition, 4*(3), 293-308.

Midgley, C., Kaplan, A., & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology., 93*, 77-86.

Mikulincer, M., Glaubman, H., Ben-Artzi, E., & Grossman, S. (1991). The cognitive specificity of learned helplessness and depression deficits: The role of self-focused cognitions. *Anxiety Research, 3*, 273-290.

Mishra, P. (2006). Affective feedback from computers and its effect on perceived ability and affect: A test of the computers as social actors hypothesis. *Journal of Educational Multimedia and Hypermedia, 15*(1), 107-131.

Mishra, P., Nicholson, M., & Wojcikiewicz, S. (2001). Does my wordprocessor have a personality? Topffer's Law and Educational Technology. *Journal of Adolescent and Adult Literacy, 44*(7), 634-641.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*(6), 669-678.

Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology, 29*(5), 1093-1110.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27*(10), 864-876.

Nease, A., Mudgett, B. O., & Quinones, M. A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology, 84*(5), 806-814.

Nicholls. (1984). Conceptions of ability and achievement motivation. In R. E. Ames & C. Ames (Eds.), *Research on motivation in education* (pp. 39-73). Orlando, FL: Academic Press.

Nitko, A. J., & Brookhart, S. M. (2007). *Educational Assessment of Students* (5 ed.). Upper Saddle River: Pearson Prentice Hall.

O'Leary, K. D., & O'Leary, S. G. (1977). *Classroom management: The successful use of behavior modification* (2 ed.). New York, NY: Pergamon.

Oosterhof, A. (2001). *Classroom applications of educational measurement*. Upper Saddle River, NJ: Merrill Prentice Hall.

Page, E. B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *Journal of Educational Psychology, 49*(2), 173-181.

Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods*. Newbury Park: Sage.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy and Practice, 5*(1).

Piaget, J. (1970/1983.). Piaget's theory. In W. Kessen (Ed.), *Handbook of child psychology* (Vol. 1: History, Theory and Methods). New York: Wiley.

Pintrich, P., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In A. Wigfield & J. S. Eccles (Eds.), *Development of Acievement Motivation*. San Diego, CA: Academic Press.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of Self-regulation*. San Diego, CA: Academic Press.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in Education: Theory, Research and Applications* (2 ed.). Upper Saddle River: Pearson Prentice Hall.

Podsakoff, P. M., & Farh, J. (1989). Effects of feedback sign and credibility on goal setting and task performance: A preliminary test of some control theory propositions. *Organizational Behavior and Human Decision Processes, 44*, 45–67.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*, 4-13.

Reeves, B., & Nass, C. I. (1996). *The Media Equation: How people treat computers, television, and new media as real people and places*. Cambridge: Cambridge University Press/CSLI.

Resnick, L. B., & Resnick, D. P. (1991). Assessing the Thinking Curriculum: New Tools for Educational Reform. In B. Gifford (Ed.), *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction.* MA: Kluwer.

Roesch, S. C. (1998). The factorial validity of trait positive affect scores: Confirmatory factor analyses of unidimensional and multidimensional models. *Educational and Psychological Measurement, 58*, 451-466.

Roos, B., & Hamilton, D. (2005). Formative assesment: a cybernetic viewpoint. *Assessment in Education, 12*(1), 7-20.

Sadler, R. (1998). Formative assessment: revisiting the territory. *Assessment in Education, 5*(1), 77-84.

Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist, 25*(1), 71-86.

Schunk, D. H. (1995). Self-efficacy and education and instruction. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment: Theory, research, and application* (pp. 281-303). New York: Plenum Press.

Scriven, M. (1967). The methodology of curriculum evaluation. In R. Taylor, R. Gagne & M. Scriven (Eds.), *AERA Monograph Series on Curriculum Evaluation* (Vol. 1, pp. 39-83). Chicago, IL: Rand McNelly.

Shanab, M. E., Peterson, D., Dargahi, S., & Deroian, P. (1981). The effects of positive and negative verbal feedback on the intrinsic motivation of male and female subjects. *Journal of Social Psychology, 115*, 195–205.

Shinn, M. R., & Hubbard, D. D. (1992). Curriculum-based measurement and problem-solving assessment-basic procedures and outcomes. *Focus On Exceptional Children, 24*(5), 1-20.

Siero, F., & Van Oudenhoven, J. P. (1995). The effects of contingent feedback on perceived control and performance. *European Journal of Psychology of Education, 10*, 13-24.

Silver, W. S., Mitchell, T. R., & Gist, M. E. (1995). Responses to successful and unsuccessful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes, 62*, 286-299.

Smith, J. K., Smith, L. F., & DeLisi, R. (2001). *Natural classroom assessment: Designing seemless instruction & assessment*. Thousand Oaks. CA: Corwin Press.

Spencer, S. (2005). Stereotype Threat in Mathematics in Undergraduate Women. Unpublished Doctoral dissertation. Rutgers University.

Stiggins, R. J. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan, 86*, 22-27.

Stiggins, R. J. (2005). *Student-Involved Assessment FOR Learning* (4 ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Stipek, D. (2002). *Motivation to learn: Integrating theory and practice* (4th ed.). Boston, MA: Allyn & Bacon.

Szalma, J. L. (2006). Training for vigilance: The effect of knowledge of results format and dispositional optimism and pessimism on performance and stress. *British Journal of Psychology, 97*, 115-135.

Szalma, J. L., Hancock, P. A., Warm, J. S., Dember, W. N., & Parsons, K. S. (in press). Training for vigilance: Using predictive power to evaluate feedback effectiveness. *Human Factors*.

Thomas, J. W. (1993). Promoting independent learning in the middle grades-the role of instructional support practices. *Elementary School Journal, 93*, 575-591.

Thorndike, E. L. (1927). The Law of Effect. *American Journal of Psychology, 39*, 212-222.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

Torrance, H. (1993). Formative assessment: Some theoretical problems and empirical questions. *Cambridge Journal of Education, 23*(3).

Tubbs, M. E., Boehne, D. M., & Dahl, J. G. (1993). Expectancy, valence, and motivational force functions in goal-setting research: An empirical test. *Journal of Applied Psychology, 78*, 361-373.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: a typology. *British Educational Research Journal, 22*, 389-404.

Vallacher, R. R., & Wenger, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94*, 3-15.

Vancouver, J. B. (2005). The depth of history and explanation as benefit and bane for psychological control theories. *Journal of Applied Psychology, 90*(1), 38-52.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527-579). Cambridge, MA: MIT Press.

Vygotsky, L. S. (1987). *The Collected Works of L. S. Vygotsky.* (Vol. 1: The Problem of General Psychology). New York: Plenum Press.

Ward, J. R., Rogers, D. A., Byrne, Z. S., & Masterson, S. S. (2004). *State Versus Trait Goal Orientation: Is There Truly A Difference?* Paper presented at the The 19th annual conference of the Society of Industrial and Organizational Psychology in Chicago, Illinois on April 2-4, 2004.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 47*, 1063–1070.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations and psychobiological evidence. *Journal of Personality and Social Psychology, 76*, 820–838.

Wolf, L. F. (1993). The effects of motivation and anxiety on test performance. *Dissertation Abstracts International, 54, 04A (University Microfilms No. 93-24654).*

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*, 227-242.

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education, 8*, 341-351.

Wolters, C. A. (2004). Advancing achievement goals theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology, 96*, 236-250.

Wolters, C. A., Yu, S., & Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences, 11*, 281-299.

Wong, B. Y. L. (1985). Self-questioning instructional research. *Review of Educational Research, 55*, 227-268.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal-setting. *American Educational Research Journal, 29*, 663-676.

APPENDICES

APPENDIX A

General Rubric for the Essay Writing Task

The following criteria will be used to evaluate your work:

1. Is the essay a cohesive, synthesized piece that centers around the subject of motivation?

2. Are all the theories properly applied, well-explained and substantiated with examples and information from class material?

3. Does the essay show you have a grasp of writing technique and mechanics?

- Are the ideas well-developed?

- Is the essay well organized with clear transitions?

- Does the essay contain errors in grammar and conventions?

- Is there variety in word choices?

4. Does the essay indicate independent interpretation, that is, is it innovative, creative, and thoughtful?

The maximum score is 100.

APPENDIX B

Demographic Questionnaire

1. What is your gender?

2. What is your age?

3. What is your country of origin?

4. What is your native language?

5. What is the highest level of education you plan to attain?

6. What is your race/ethnicity?

APPENDIX C

Performance and Learning Goal Orientation Items

Performance goal orientation

1. I prefer to do things that I can do well rather than things that I do poorly.

2. I'm happiest at work when I perform tasks on which I know that I won't make any errors.

3. The things I enjoy the most are the things I do the best.

4. The opinions others have about how well I can do certain things are important to me.

5. I feel smart when I do something without making any mistakes.

6. I like to be fairly confident that I can successfully perform a task before I attempt it.

7. I like to work on tasks that I have done well on in the past.

8. I feel smart when I can do something better than most other people.

9. Even if I know that I did a good job on something, I'm satisfied only if others recognize my accomplishments.

10. It's important to impress others by doing a good job.

Learning goal orientation

1. The opportunity to do challenging work is important to me.

2. When I fail to complete a difficult task, I plan to try harder the next time I work on it.

3. I prefer to work on tasks that force me to learn new things.

4. The opportunity to learn new things is important to me.

5. I do my best when I'm working on a fairly difficult task.

6. I try hard to improve on my past performance.

7. The opportunity to extend the range of my abilities is important to me.

8. When I have difficulty solving a problem, I enjoy trying different approaches to see which one will work.

9. On most jobs, people can pretty much accomplish whatever they set out to accomplish.

10. Your performance on most tasks or jobs increases with the amount of effort you put into them.

APPENDIX D

Post-Test Motivation Scale Items

1. Doing well on this exam was important to me.

2. I am concerned about the grade I receive on this exam.

3. This was a very important exam to me.

4. I gave my very best effort on this exam.

5. I could have worked harder on this exam.

6. I did not give this exam my full attention.

7. I am eager to find out how well I did on this exam.

8. I was highly motivated to do well on this exam.

APPENDIX E

Self-Efficacy Scale Items

1. This is a subject area that I usually do well in.

2. I have skills and knowledge necessary to have succeeded on this exam.

3. I am confident I received a high grade on this exam.

4. I successfully applied lecture and readings content to answer the exam question.

5. I am capable of doing well on this kind of exams.

6. I doubt that I did well on this exam.

7. I did as well or better than most other students on this exam.

8. I am not competent enough to have done well on this exam.

APPENDIX F

The Positive and Negative Affect Scale

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to the word. Indicate to what extent you feel like this right now, that is, at the present moment. Use the following scale to record your answers.

1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

_____ interested   _____ irritable

_____ distressed   _____ alert

_____ excited   _____ ashamed

_____ upset   _____ inspired

_____ strong   _____ nervous

_____ guilty   _____ determined

_____ scared   _____ attentive

_____ hostile   _____ jittery

_____ enthusiastic   _____ active

_____ proud   _____ afraid

APPENDIX G

Rubric for Grading the Content of an Essay

Table G1

Content Grading Rubric

Score	# of Theories	Criteria for Evaluation
0	0	No content (word "motivation" doesn't count)
1	0	Several relevant terms, not explained or used inappropriately
1.5	1	One or two theories mentioned appropriately, but the description is not full or confused
2	1	One theory explained, other terms are used inappropriately or too lightly
2.5	1	One theory well-explained, others are touched upon correctly (terms mentioned)
3	2	Two theories explained, but with some confused application, not enough detail and examples. (some other theories may be touched on)
3.5	2	Two theories explained, description of one not full/confused (some other theories may be touched upon)
4	2	Two theories well-explained, and/or terms from one or more theories mentioned
4.5	2	Level 4 plus argument leading very well to conclusion
5	3+	Three or more theories explained and properly applied, but with some confused terms and not enough detail for one of them
5.5	3+	Three or more discussed theories, well-explained and properly applied, with minor omissions
6	3+	Three or more discussed theories, well-explained, properly applied and substantiated by examples; other class readings are included

APPENDIX H

Experimental Conditions

Table H1

Groups Formed by Factor Crossings

	No grade		Grade	
	No praise	Praise	No praise	Praise
No feedback	No feedback	No feedback	No feedback	No feedback
	No grade	No grade	Grade	Grade
	No praise	Praise	No praise	Praise
Computer feedback	Computer feedback	Computer feedback	Computer feedback	Computer feedback
	No grade	No grade	Grade	Grade
	No praise	Praise	No praise	Praise
Instructor feedback	Instructor Feedback	Instructor feedback	Instructor feedback	Instructor feedback
	No grade	No grade	Grade	Grade
	No praise	Praise	No praise	Praise

CURRICULUM VITA

Anastasiya A. Lipnevich

Education:

RUTGERS UNIVERSITY

Ph.D., Educational Psychology, 2007

Ed.M., Educational Psychology, 2004

BELARUSIAN STATE PEDAGOGICAL UNIVERSITY

M. S., Psychology and Education, combined degree, 2002

Equivalent to a US M.S., with Bachelor's en passant

M. S., Italian Language, 2000

Equivalent to a US M.S., with Bachelor's en passant

Professional Experience:

2004–2007   RUTGERS UNIVERSITY

Rutgers, the State University in New Jersey

Graduate assistant; Research program director

2006 Jan–2007   KEAN UNIVERSITY

Adjunct instructor of psychological statistics, general psychology

2002 Jan–May   BELARUSIAN STATE PEDAGOGICAL UNIVERSITY

Instructor of educational psychology.

2000–2001   BELARUSIAN STATE PEDAGOGICAL UNIVERSITY

Instructor, Italian language.

1997–2001   COMITATO CHERNOBYL FRIULI-VENEZIA-GIULIA

Counselor, Italian–Russian translator

Instructor, Italian language.

Publications:

Lipnevich, A. A., Beder, H., Robinson-Gellar, P. (in press). Typology of adult literacy education. *Adult Basic Education*.

Lipnevich, A. A., Beder, H. (in press). Adult literacy education students and PhD students: Two ends of the self-esteem Continuum?. *Adult Basic Education*.

Lipnevich, A. A. (2006). Self-esteem: Myth or reality. *Focus on basics*, 8 (2), 1-5.

Robinson-Gellar, P., Lipnevich, A. A. (2006). Instructional practices of ABE and GED teachers. *Focus on basics*, 8 (2).

Conference Presentations:

Lipnevich, A. A., Smith, J. K. (2007, April). Russian and American perspectives on self-regulated learning. Presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Valentine, T., Lipnevich, A., Milton, J., Beder, H. (2006, May). *Mapping cognitive engagement in adult literacy classrooms.* Presented at the annual meeting of the Adult Education Research Conference, Minneapolis, MN.

Beder, H., Lipnevich, A., Valentine, T., Milton, J. (2006, April). *Outcomes of cognitive engagement.* Presented at the annual meeting of the Commission on Adult Basic Education, Houston, TX.

Lipnevich, A., Beder, H. (2006, April). *Reevaluating the premises of the deficit model for adult education.* Presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Lipnevich, A. (2005, November). *Self-esteem: Logical but unexpected findings.* Presented at the annual meeting of the American Association for Adult and Continuing Education, Pittsburgh, PA.

Valentine, T., Lipnevich, A., Milton, J., Beder, H. (2005, November). *Working hard on learning: Cognitive engagement in the adult literacy classroom.* Presented at the annual meeting of the American Association for Adult and Continuing Education, Pittsburgh, PA.

Valentine, T., Lipnevich, A., Milton, J., Beder, H. (2005, June). *The role of cognitive engagement in adult literacy learning.* Presented at the Adult Education Research Conference, Athens, GA.

Lipnevich, A. (2005, May). *Self-esteem: Considerations for adult literacy education.* Presented at the annual meeting of the Commission on Adult Basic Education, Anaheim, CA.

Beder, H. Lipnevich, A., Valentine, T., Milton, J. (2005, May). *Engagement in instruction.* Presented at the annual meeting of the Commission on Adult Basic Education, Anaheim, CA.