J Behav Educ (2009) 18:245–265
DOI 10.1007/s10864-009-9089-2

ORIGINAL PAPER

The Differential Effects of Elaborate Feedback and Basic Feedback on Student Performance in a Modified, Personalized System of Instruction Course

Jared A. Chase · Ramona Houmanfar

Published online: 6 August 2009

Abstract Educators in large-enrollment courses are faced with the challenge of effectively disseminating information to their students to ensure that they learn the content provided. A related issue involves the means by which instructors evaluate student performance. Offering effective forms of performance feedback may be one technique to provide students with additional information to facilitate learning. Accordingly, the purpose of this investigation was to determine the effects of elaborate feedback and basic feedback on student performance. Two groups from an introductory psychology course participated in the current study. The Basic Feedback Group ($N = 108$) received basic feedback on all quizzes. The Elaborate Feedback Group ($N = 102$) received elaborate feedback on all quizzes. Response accuracy and learning gain were evaluated between groups. Visual analyses demonstrated the relative effectiveness of elaborate feedback on subsequent student performance. Descriptive and inferential statistical analyses revealed that elaborate feedback was beneficial in general and particularly for questions that were determined to be difficult by item analyses. Results and implications are discussed in further detail.

Keywords Elaborate feedback · Basic feedback · Personalized system of instruction (PSI) · Large-enrollment course

Introduction

A prevalent concern facing educators is the question of how to effectively disseminate information to their students to ensure that they learn the content presented. A related issue involves the means by which instructors evaluate student

J. A. Chase (✉) · R. Houmanfar
Department of Psychology/296, University of Nevada, Reno, Reno, NV 89557, USA
e-mail: jchase@unr.nevada.edu

performance to address the aforementioned concern. This matter is exacerbated in large-enrollment courses in university settings. In such environments, many instructors have relied upon multiple-choice tests to evaluate student performance (Cross 1995). Oftentimes, tests and exams are graded electronically (e.g., on Scantron© forms) in which feedback regarding performance is generally deferred and in the form of a letter or a number grade. The implication is that if the original test items are not accompanied with the student's results, the student will not know specifically which questions were answered correctly or incorrectly and more importantly, why. Alternatively, many large-enrollment courses are increasingly utilizing web-based technologies to grade tests and exams which provide computerized feedback. However, this feedback is often of a basic nature (e.g., simply indicating whether the response was correct or incorrect).

Feedback is a construct that has been extensively studied in both the psychological and educational literature. Furthermore, vast amounts of research in the organizational behavior management (OBM) literature have examined the provision of feedback as it pertains to training, management, and employee performance (e.g., Alvero et al. 2001; Balcazar et al. 1985). In organizational settings, performance feedback has been defined as information presented to individuals pertaining to the quantity and/or quality of past performance (Prue and Fairbank 1981). Other definitions of performance feedback in this context include: (a) information transmitted back to the individual following a particular performance (Sulzer-Azaroff and Mayer 1991), (b) information about performance that allows an individual to adjust one's performance (Daniels 1994), and (c) information that notifies performers as to how well they are doing (Rummler and Brache 1995).

Although a consensus regarding the appropriate definition of performance feedback has yet to be made, feedback in general has been regarded as an effective and an efficient means to improve performance. For example, feedback has been implemented to improve task completion and time allocation with graduate students (Houmanfar and Hayes 1998); therapist and patient performance in psychiatric group homes (Huberman and O'brien 1999); to decrease cash register shortages (Rohn et al. 2002); to increase safe driving behaviors (Hickman and Geller 2003); in retail settings (Pampino et al. 2003); to improve customer service (Eikenhout and Austin 2005); and to increase safe ergonomic behaviors in office settings (Sasson and Austin 2005).

While the effectiveness of feedback in improving performance in organizational environments has repeatedly been demonstrated, it should be noted that feedback is often implemented as one component of a "treatment package." For example, feedback is often delivered in conjunction with goal setting procedures and incentives for desired performance. Furthermore, the effects of feedback may be confounded by many other interdependent variables. For instance, feedback may be delivered at the individual or group level; it can be written, verbal, and/or graphically displayed; and it can be mediated by the individual, by peers, or by supervisors. In addition, feedback may be delivered immediately following the target response or it may be delayed (Fairbank and Prue 1982). Each of these parameters may differentially influence performance. That is, several variables and

the context in which feedback is delivered may influence or have no effect on behavior. Accordingly, Houmanfar and Hayes (1998) contend that, "the effects of feedback vary so greatly with the mode and manner of its presentation that summary claims about its effectiveness are difficult to make" (p. 70).

In addition, the content of the feedback provided is an important variable to be considered. For example, feedback may merely indicate whether a response was performed correctly or not, or the type of feedback delivered may provide more detailed information as to why the response was correct or incorrect. Just as there is no consensus regarding the precise definition of feedback, there is little consistent agreement as to what is the most effective *type* of feedback. This may be reflective of the fact that the explicit distinction between different types of content feedback and their differential effects on performance have not been systematically evaluated within the OBM literature.

Nonetheless, Braksick (2000) has argued that feedback is one of the most powerful consequences one can deliver to shape desired performance. She further delineates the distinctions between positive, negative, and constructive feedback. Braksick (2000) proposed that positive feedback is used to encourage desired behavior whereas negative feedback focuses on the "bad side" and does little to improve performance (p. 146). In this sense, negative feedback is likened to a form of criticism and is often directed at the individual rather than to the undesired *behavior* of the individual. As an alternative to negative feedback, Braksick (2000) suggests the use of *constructive feedback*. She asserts that constructive feedback is intended to, "discourage an undesired behavior and replace it with a preferred behavior" (p. 146). Constructive feedback both punishes (i.e., reduces the probability of) the undesired behavior by describing what was wrong with it and further, specifies what the preferred behavior would be. In short, constructive feedback includes pinpointing the behavior we wish to discourage, while also pinpointing preferred behaviors. The implication here with respect to education is apparent; instead of providing aversive forms of feedback to correct student performance, feedback can be both positive and instructive. However, only a few studies have made the distinction between feedback *types* with respect to performance in educational settings (e.g., Arnett 1985; Bumgarner 1984; Farragher and Szabo 1986; Heald 1970; Pressey 1926; Sassenrath and Gaverick 1965).

Nevertheless, the different types of feedback offered to students can generally be classified into one of four categories. One type of feedback (e.g., right/wrong) simply informs students as to whether their responses were correct or incorrect (Arnett 1985; Bumgarner 1984). In this case, no other information indicating what the correct answer would be is provided if the test item was answered incorrectly. A second type of feedback (e.g., corrective) indicates what the correct answer option would be if answered incorrectly (Sassenrath and Gaverick 1965). Likewise, Bangert-Drowns et al. (1991) suggested that, "effective feedback should not only signal the correctness or incorrectness of an answer; it should also provide correction when necessary" (p. 228). Another form of feedback allows students to repeatedly answer questions until all are answered correctly (Heald 1970; Pressey 1926). A fourth type of feedback (e.g., explanatory) provides elaboration as to *why* particular question items were answered correctly or not (Farragher and Szabo

1986). This fourth type of feedback is comparable to Braksick's (2000) *constructive feedback* and will be defined as *elaborate* feedback henceforth (see "Appendix 1"). In addition, elaborate feedback is contrasted with corrective feedback which is more *basic* in nature (e.g., correct or incorrect). The limited research on feedback *type* necessitates further investigation before any conclusions regarding the efficacy of each can be made.

Of note, the systematic study of feedback by experimental psychologists can be traced to Thorndike's (1913) *law of effect* (Vollmeyer and Rheinberg 2005). Thorndike's famous puzzle box experiments led him to conclude that environmental feedback in the form of satisfying consequences as a result of a behavioral response will influence the probability of the organism engaging in similar behaviors in the future under similar conditions. In this sense, environmental feedback promotes learning. As such, the feedback construct as it pertains to education has received considerable attention.

In line with Thorndike's perspective, Skinner promoted a shift in educational practices from that of aversive control to the practice of reinforcing appropriate behaviors and accurate responses. Skinner's (1938) definition of reinforcement closely parallels Thorndike's *law of effect* in that the probability of a response occurring in the future is increased when that response has been previously consequated by a reinforcing event. As such, Skinner developed a *teaching machine* which systematically incorporated the principles of operant conditioning to the educational environment. Skinner's teaching machine provided the individual with immediate feedback in which correct responses were reinforced which further promoted the discriminations between correct and incorrect responses. Moreover, Skinner (1968) contended that, "The application of operant conditioning to education is simple and direct. Teaching is the arrangement of contingencies of reinforcement under which students learn..." (pp. 64–65). From Skinner's perspective, the role of an instructor evolves from that of controlling student behavior and performance through aversive control to one of arranging special contingencies which facilitate learning. The essential premise is that *the learner knows best*. That is, if a given student is not meeting performance standards, the problem may not reside within the individual, but with the instructional contingencies which have not acquired stimulus control over the behavioral repertoire of accurate responding to instructional antecedents.

Personalized System of Instruction

In line with Skinner, Fred Keller concluded that a majority of educational environments employ aversive contingencies to manage student performance and behavior. As such, Keller (1968) introduced the behavior analytic community to a personalized system of instruction (PSI) in which special contingencies are arranged to facilitate learning in large-enrollment courses which further omitted aversive contingencies. PSI was an application of reinforcement theory. The five formal components of PSI include: (a) self-pacing, (b) a unit-mastery requirement, (c) the use of lectures for motivational purposes, (d) an emphasis upon the written word,

and (e) the use of undergraduate proctors. In Keller's system, proctors allowed for repeated testing, immediate scoring, and served a social function. More importantly, proctors offered tutoring and provided students with feedback regarding performance.

The effects of PSI were immediately apparent. Specifically, the grade distribution of Keller's students was negatively skewed with a greater proportion of students receiving A's and B's in the course. This is contrasted with conventional large-enrollment courses in which a normal distribution is typically observed. Since 1968, many variations of PSI-type courses have been implemented and evaluated in which the results obtained by Keller have been replicated (Austin 2000; Buskist et al. 1991; Kulik et al. 1980; Sherman 1992). In particular, Kulik et al (1980) evaluated the effectiveness of five types of instructional technologies and college teaching. The five types of instructional technologies included (a) PSI, (b) computer-based instruction, (c) programmed instruction, (d) Postlethwait's Audio-Tutorial approach, and (e) visual-based instruction. Among all variables associated with college teaching that were evaluated, results suggested that the *type* of technology implemented was the most influential variable on performance. Specifically, the authors proposed that PSI-type courses had the strongest effect on student achievement and the most pronounced effects on student ratings of instruction (Kulik et al. 1980, p. 204). The authors further concluded that immediate feedback on student performance is one of the variables responsible for the effectiveness of PSI.

Purpose

Although the immediacy with which feedback delivered to students has been examined at length (see Kulhavy 1977; Kulik and Kulik 1988), empirically based examinations of different *types* of feedback and their effects on student performance in PSI-type courses is conspicuously lacking from the relevant literature. It is assumed that qualitatively different types of feedback may yield differential results with respect to student performance in this type of an educational environment. Furthermore, Austin (2000) has argued that instructors should incorporate instructional methods that are effective, efficient, and empirically validated. Research has demonstrated the relative effectiveness of PSI. Moreover, the structure of PSI also allows for the continual self-examination of procedures and component parts that enhance the learning process. Similarly, feedback (a component of PSI) has emerged in the literature as a means to facilitate both student learning and as a means to enhance teaching performance. As such, the specific *type* of feedback delivered to students in a PSI course warrants further investigation.

Accordingly, the purpose of this investigation was to determine the differential effects of *elaborate* feedback and *basic* feedback on student performance in a large-enrollment introductory psychology course. Students in a PSI-type course at a northwestern university in Nevada have historically received electronic feedback regarding their performance immediately after the submission of quizzes and exams during the last 10 years with corresponding stable grade distributions. However, this feedback has traditionally been of a basic nature (i.e., correct or incorrect). As such,

the question this study endeavored to answer was the following: "Would providing students with a more elaborate form of feedback have any impact on their performance in a statistically and/or practically significant way?"

Methods

Participants and Setting

Students from two of five sections ($N = 525$) of an introductory psychology course participated in the current investigation. These participants were primarily freshman students ($\sim$18–19 years of age), from diverse racial backgrounds, with an approximately equal gender split. Prior to the study, each section was randomly selected and assigned to experimental or control conditions. The experimental condition consisted of providing participants in one section ($N = 102$) with electronic, elaborate feedback following the submission of each quiz and exam. The control condition was comparable to the experimental condition with the exception that only basic feedback was provided to participants in this section ($N = 108$).

Participants completed all quizzes and exams in a laboratory designated for an introductory psychology course. The quiz room was $\sim$20 m $\times$ 10 m in which 14 Dell® Pentium IV personal computers were provided to the participants. Each computer was connected to the Internet. WebCT® (a web-based instructional shell) graded all quizzes and exams which also provided participants with immediate performance feedback.

Experimental Design

Between-group Analysis

A between-group design was implemented in the current study. The Basic Feedback Group (BF Group; $N = 108$) received only basic feedback on all quiz and exam questions. The Elaborate Feedback Group (EF Group; $N = 102$) received elaborate feedback on all quiz and exam questions.

Procedures

All course content was presented by six instructors in the same order and within the same time frame (e.g., one chapter per week) for all participants. All instructors followed a standard protocol for all discussion sessions in that students could choose to attend discussion groups from any or all instructors over the course of the semester. While participants could choose among various discussion groups to attend, the internal validity of this study was maintained in that only the primary independent variable (i.e., feedback type) was manipulated through the automated testing system while instructors followed a standardized protocol for all discussion groups. The course required the completion of ten, 15 multiple-choice question

quizzes for each of the assigned chapters in addition to a 60-question midterm and a 120-question final exam.

Quiz and Exam Question Selection

The WebCT® test bank utilized in this study contained ~70–100 questions per chapter and was programmed to randomly select 14 questions for each chapter quiz. This yielded a total of 140 quiz questions. Once these questions were randomly selected, elaborate feedback was programmed for each. The elaborate feedback manipulation was deleted from each of these quiz questions for the BF Group.

A and B Versions for Each Chapter Quiz

The structure of the instructional system (i.e., PSI) used in this study allowed for participants to re-take chapter quizzes to acquire mastery. To reduce the potential of practice effects, two versions of each chapter quiz were created. *Version A* consisted of the randomly selected experimental questions that would be presented a second time (e.g., on the mid-term or the final). *Version B* comprised the remaining four quiz re-take opportunities for each chapter quiz. This version consisted of 14 randomly selected quiz questions (from the entire test bank) with only basic feedback (i.e., 100 or 0% correct) provided. None of the questions from *Version A* were repeated again in any of the *B* versions. To ensure that participants came into contact with the experimental question items, a selective quiz release was programmed into WebCT® in which participants could not take *Version B* quizzes unless the *Version A* quiz for a given chapter had been completed. Further, *Version A* quizzes for each chapter could only be taken one time. That is, none of the quiz questions used in the present analysis were presented a second time on any of the chapter quiz re-take opportunities.

WebCT® was then programmed to randomly select questions from the *Version A* quizzes to be presented on the mid-term and final to assess learning gain on the participants' second attempt. Once selected, these questions were held constant across each group. In particular, identical questions delivered in the same order were presented to each of the two groups on all *Version A* quizzes and exams for the purpose of this investigation. Only the feedback component was manipulated.

Independent Variable and Operational Definitions

Elaborate Feedback

Elaborate feedback was defined as written, verbal feedback delivered electronically immediately after the submission of each quiz. Elaborate feedback was further distinguished by two components: (a) *Specific feedback* is specific to each question and includes information indicating which answer options are correct and incorrect, and (b) *General feedback* which includes conceptual information, definitions, as well as referential page numbers from the text ("Appendix 1").

Basic Feedback

Basic feedback was defined as numeric feedback (i.e., 100 or 0%) as to whether participants answered a given question correctly. In addition, following any incorrect answer, the correct response was indicated ("Appendix 2"). Basic feedback was used as the control condition in this study as this is the type of feedback that has been historically delivered to participants since the inception of this instructional system and automated testing.

Dependent Variables: Primary Measures

Accuracy of Response

Response accuracy to quiz items and questions from the mid-term and final were evaluated and compared between groups. Specifically, accuracy of responding was evaluated two times: (a) on the participants' first attempt on each *Version A* quiz and (b) on the mid-term or final exam. The dependent measure of response accuracy was calculated by summing the correct responses by all participants and dividing that number by all possible quiz questions delivered to yield an aggregate percentage score. The current design allowed for the evaluation of response accuracy on 140 quiz question responses, 60 mid-term responses, and 60 final exam responses for each participant and across groups. Analyses of responses to specific questions and overall scores were conducted to determine the differential effects of both types of feedback.

Learning Gain

Learning gain was evaluated by examining response accuracy on participants' first attempt (i.e., on *Version A* quiz questions) and comparing response accuracy to those same questions on their second attempt (i.e., on either the mid-term or the final exam). In particular, 60 (12 per chapter) of the 70 questions from the first, five *Version A* quizzes were randomly selected and presented on the midterm. In addition, 60 (12 per chapter) of the 70 questions from the last five *Version A* quizzes were presented on the final exam. The responses on the participants' second attempt (for both the mid-term and the final exam) were compared to performances on their first attempt.

Performance on Difficult Questions

Item analyses were conducted in the current investigation to determine the difficulty of each question and further, to identify whether there was an interaction effect between the type of feedback delivered and the relative difficulty of each question. Item analysis procedures provide information such as the *difficulty* of each item, the discrimination power of the item, and other properties of choices or distracters (Crocker and Algina 1986; Henrysson 1971). The top quartile of all possible questions that were answered incorrectly on participants' first attempt was defined

as the "most-difficult questions". This yielded 30 questions for the BF Group and the EF Group, respectively. Accuracy on first and second attempts for these questions as well as overall learning gain was evaluated.

Dependent Variables: Secondary Measures

Grade Distribution

A grade distribution for each of the two groups was evaluated. Only quiz and exam scores were included in this analysis. For example, all scores from any other assignments and extra credit were excluded from the present grade distribution analysis.

For the purpose of comparison, an analysis of comparable semesters (fall 2003 through 2005) was conducted. For these analyses, all sections, all scores including quiz and exam totals, assignments, and extra credit were included to achieve consistent reliability with previous grade distribution analyses which are regularly conducted as part of the evaluation of this instructional system.

Course Survey

At the end of each semester, the instructional system used in this study distributes course evaluations inclusive of 18 questions to identify global levels of satisfaction with the class. Specific to the current investigation, two additional "yes/no" questions were included in the course survey to identify relative levels of satisfaction as they directly relate to the feedback manipulation. For instance, participants receiving elaborate feedback were asked whether the individual had noticed that feedback was provided after each question. In addition, a follow-up question asked if they found elaborate feedback to be helpful in preparing them for future exams in this course.

Data Collection and Analysis

WebCT® stored all data electronically for each participant and each question. All data were exported from WebCT® to Microsoft® Excel workbooks for the purpose of graphic, visual analyses. Specifically, data from each quiz and exam were exported into separate Microsoft® Excel worksheets. Accuracy of answers to questions on the mid-term and final exam was compared to accuracy on previous quizzes. Learning gain was assessed by subtracting mean performances on questions for the participants' first attempt from the mean performance scores on these questions on subsequent attempts.

Visual analyses allowed for the identification of differences between groups affected by the independent variable (Sidman 1960). Basic descriptive statistical analyses were conducted to identify mean differences between groups. In addition, relevant data were exported from the Microsoft® Excel workbooks into SPSS® software for inferential statistical analyses. Specifically, factorial analyses of variance (ANOVAs) were conducted to divide scores on quizzes and exams into

systematic sources of variability that reflected the experimental manipulation from unsystematic sources that were not influenced by the independent variable. Additionally, *eta squared* was calculated to determine the effect size of the present intervention on primary measures and probabilities were calculated using Fisher's exact F-probability procedure.

Interobserver Agreement

As previously mentioned, WebCT® stored all data electronically. Therefore, interobserver agreement (IOA) was obtained in the following manner: The first author exported all electronic data from WebCT® into a Microsoft® Excel workbook. A trained undergraduate student followed the same procedure by exporting the same data into a separate Microsoft® Excel workbook. True/false tests were then conducted for each cell containing data to determine a correspondence between each data set. No disagreements were observed and 100% reliability was obtained. These data were then used for subsequent statistical and graphic analyses.

Results

The results of the between-group analysis on primary and secondary measures will be discussed below. In addition, both graphic depiction and the results of statistical analyses will be presented.

Primary Measures

Accuracy of Response

Inspection of the data revealed that performance was undifferentiated between the two groups on the participants' first attempt. Specifically, descriptive statistics (i.e., measures of central tendency) were utilized to determine the mean accuracy of responding on the first attempt. When all results were aggregated, accuracy of responding was 69% for both groups. In addition, there was improvement from the first attempt to the second attempt across both groups. However, the relative level of improvement on the second attempt was greater in the elaborate feedback condition when compared to the basic feedback condition (Table 1).

Learning Gain

It should be emphasized that neither group received either type of feedback until the submission of their quizzes on their first attempt. Therefore, learning gain per se was assessed on the second attempt (i.e., the mid-term exam for questions from the first, five chapter quizzes and the final exam for questions from the last, five chapter quizzes). As with the previous analysis, accuracy of responding was determined on the participants' second attempt using measures of central tendency to determine the mean percentage correct on each question as well as the standard deviation

Table 1 Mean percentage correct and standard deviations for each type of feedback

	M (1st attempt) (%)	SD	*M* (2nd attempt) (%)	SD
All questions				
Basic feedback	69	.141397	80	.120706
Elaborate feedback	69	.148706	83	.104217
30 Easiest questions				
Basic feedback	87	.042614	91	.05394
Elaborate feedback	87	.036070	93	.04728
30 Hardest questions				
Basic feedback	50	.073439	64	.090323
Elaborate feedback	49	.085129	70	.095088

Mean percentage correct and standard deviations on all questions, easy questions, and hard questions on first and second attempts with basic or elaborate feedback

(Table 1). The mean percentage correct on the second attempt was compared to accurate responding on the first attempt. Visual inspection of the data revealed that performance on the first attempt varied across groups and specific quizzes. As such, it was important to investigate the relative level of learning gain across both groups before any meaningful comparisons could be made.

Figure 1 depicts the relative percentage learning gain on the second attempt on questions that received both types of feedback across groups. Results suggested that learning gain for participants receiving elaborate feedback (the EF Group) was greater for questions from eight of the ten chapter quizzes and equal to basic feedback on questions that came from quiz 6. Interestingly, although learning gain

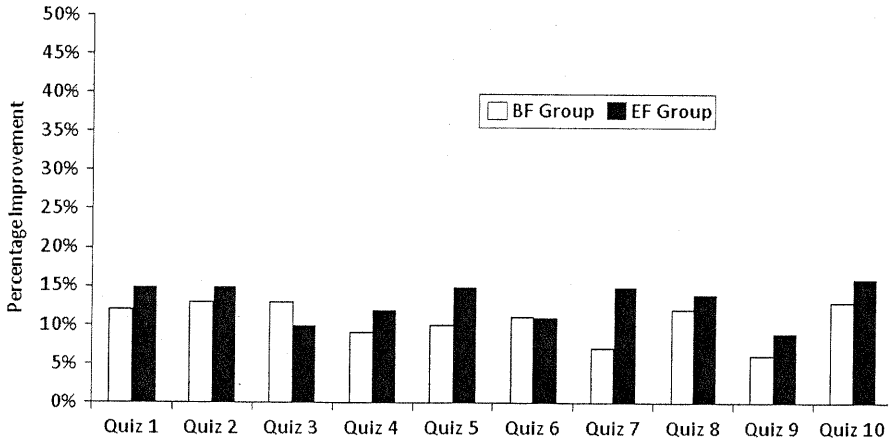

Fig. 1 Relative percentage learning gain on all questions for the Basic Feedback and Elaborate Feedback Groups. The mid-term represents the second attempt on questions from the first, five chapter quizzes. The final exam was the second attempt on questions from quizzes six through ten

was observed across both conditions, basic feedback on questions from one quiz (i.e., quiz #3) produced a greater percentage learning gain relative to elaborate feedback on the second attempt.

As previously mentioned, the percentage of all questions answered correctly on the first attempt across both groups was 69%. The percentage of those questions answered correctly on the second attempt was 80 and 83% for the BF Group and for the EF Group, respectively (Table 1). The percentage improvement (learning gain) for the BF Group was ~11 and 14% for the EF Group. This represents an overall 3% difference in learning gain between the two groups.

However, statistical analyses demonstrated that the effect of elaborate feedback relative to basic feedback was statistically significant. Results from a 2-factor analysis of variance (ANOVA) revealed that elaborate feedback (on the 120 experimental questions) resulted in a significant improvement relative to basic feedback across the BF Group and the EF Group, $F(1, 236) = 12.13$, $p < .01$ (Table 2). The probability of a Type I error was maintained at .05 for all subsequent analyses.

Question Difficulty

It was observed that performance improved on participants' second attempt across both groups regardless of question difficulty suggesting possible practice effects. Therefore, statistical tests were conducted to assess the relative effects of both basic and elaborate feedback with respect to the *easier* and the *more difficult* questions as determined by item analyses. Overall, there was a significant effect of question difficulty between groups, 7.9 vs. 15.9%; ($M = .16$, SD = .07), $F(1,236) = 106.71$, $p < .01$ (Table 2). As may be expected from the results of the previous statistical analysis comparing basic feedback to elaborate feedback, a 2-factor ANOVA revealed that there was a significant interaction between factors (i.e., feedback type and question difficulty) with the simple effect of elaborate feedback on difficult questions resulting in greater performance on those questions that received basic feedback 8.0 vs. 18.7%; ($M = .13$, SD = .08), $F(1,236) = 14.53$, $p < .01$ (Table 2).

Table 2 Analysis of variance comparing basic feedback to elaborate feedback on learning gain

Source	*df*	*F*	η^2	*p*
Between groups				
Feedback type (A)	1	12.13**	.049	.000591
Question difficulty (B)	1	106.71**	.311	.000001
A × B	1	14.53**	.058	.000176
B within-group error	236	(.003602)		

Two-factor analysis of variance comparing the effects of elaborate feedback and basic feedback on student performance. Learning gain was determined by comparing performance on first and second attempts. Values enclosed in parentheses represent mean square errors. ** $p < .01$

Performance on the Most-difficult Questions

Additional statistical analyses were conducted to evaluate the simple effects of basic and elaborate feedback with respect to the 30 most-difficult questions as indicated by performance on participants' first attempt. Results from an ANOVA revealed that exposure to the elaborate feedback manipulation resulted in subsequent improved performance on the most-difficult questions ($M = .08$, $SD = .05$), $F(1,58) = 15.21$, $p < .001$ (Table 3). For the BF Group, performance improved from 50% of these questions answered correctly on the first attempt to 64% on the second attempt. Alternatively, the percentage increase for the EF Group improved from 49% on the first attempt to 70% on the second attempt, or an overall 21% improvement (Table 1). The change in mean score by feedback type and question difficulty is further illustrated in Fig. 2.

Secondary Measures

Grade Distribution Analysis

This distribution was inclusive of only quiz and exam scores. Analyses demonstrated that the BF Group received the lowest percentage of A's and B's as

Table 3 Analysis of variance on the 30 most-difficult questions

Source	*df*	*F*	η^2	*p*
Between groups				
Between groups	1	15.21**	.208	.000252
Within-group error	58	(.004588)		

Analysis of variance comparing the effects of elaborate feedback and basic feedback on the 30 most-difficult questions. Values enclosed in parentheses represent mean square errors. ** $p < .01$

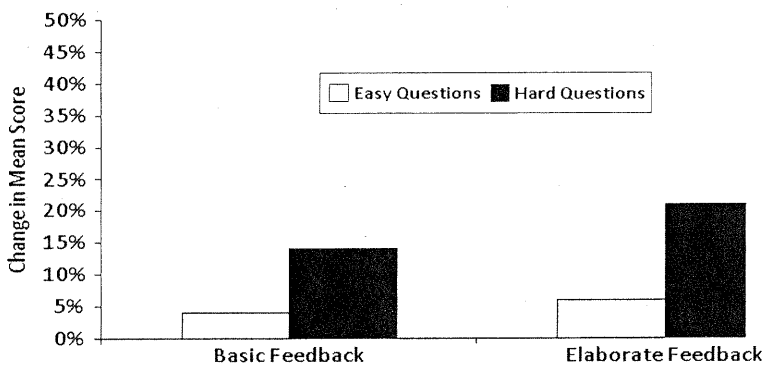

Fig. 2 Interaction effect of feedback type and question difficulty represented by a change in mean score from first to second attempts

Table 4 Overall grade distribution analysis

	A (%)	B (%)	C (%)	D (%)	F (%)
BF Group (basic feedback)	17	32	27	12	11
EF Group (elaborate feedback)	29	41	12	9	9

Grade distribution for the Basic Feedback Group and the Elaborate Feedback Group. The percentage of participants in each group receiving grades of an A, B, C, D, or an F in the course is indicated in the corresponding cells for each group

well as the highest percentage of C's, D's, and F's. This is contrasted with the EF Group which attained the highest percentage of A's and B's and the lowest percentage of C's, D's, and F's. Table 4 depicts the overall grade distribution for both groups. This grade distribution was broken down further to identify the percentage of participants who received a "B" or greater in the course as well as a "C" or less. Approximately half (49%) of the participants in the BF Group received a B or greater in the course while the other half (51%) received a C or less. Upon inspection of the data, a clear distinction between the BF Group and the EF Group became evident. That is, 70% of participants in the EF Group received a B or greater with the other 30% of participants in this group receiving a C or less in the course.

Figure 3 presents the grade distribution for comparable fall semesters immediately prior to and after this study. The present investigation was conducted during the fall 2004 semester. As demonstrated in Fig. 3, an appreciable shift in the percentage of A's was realized this semester. Interestingly, during the subsequent fall semester, the experimental manipulation was withdrawn with a corresponding reduction in the percentage of A's attained in the course. That is, the grade

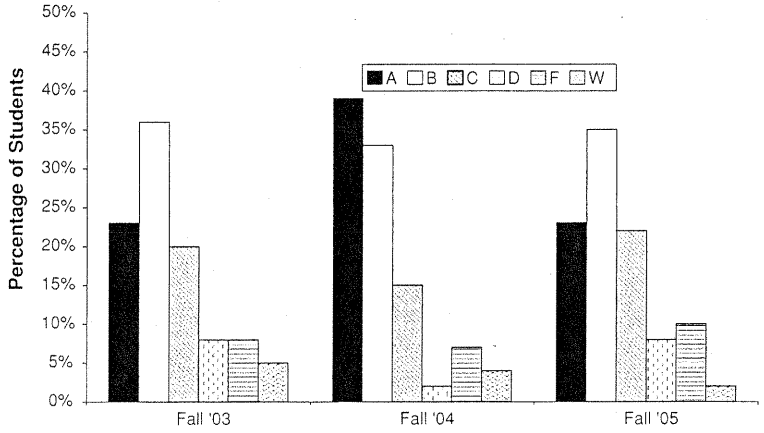

Fig. 3 Historical grade distribution comparing fall semesters from 2003 through 2005. The present study was conducted during the fall 2004 semester

distribution stabilized and closely paralleled previous semesters that did not come into contact with the elaborate feedback manipulation. Although not explicitly planned a priori, experimental control in the form of a natural reversal was demonstrated in which the grade distribution for the fall 2005 semester stabilized to "baseline levels" when treatment was withdrawn.

Course Survey

Wolf (1978) emphasized that no matter how effective our treatments and interventions appear, they will not be accepted, utilized, and maintained by our consumers if they are not satisfied with both the results of treatment and the intervention procedure itself. Accordingly, a course survey inclusive of 18 questions was provided to all students at the end of the semester. Two questions from the survey are particularly indicative of student satisfaction with our instructional system (e.g., PSI). One question inquired, "Would you recommend this course to a friend?" Eighty-two percent of respondents answered in the affirmative. The other question asked specifically, "If given the choice, would you choose to learn this way in other courses?" Sixty-seven percent of the participants who responded to the survey answered "yes".

Two questions were added to the course survey to glean information as it pertained to the feedback manipulation in this study. The EF Group ($N = 102$) was asked whether the individual had noticed that feedback was provided after each question. Sixty-eight participants responded to this question. Ninety percent ($N = 62$) answered that they had noticed the elaborate feedback manipulation. As a follow-up to this question, another was added that queried, "If your answer to the previous question was yes, did you find this type of feedback helpful in preparing you for future exams in this course?" Sixty-four participants responded to this question of which 95% ($N = 61$) of respondents agreed that elaborate feedback helped them prepare for subsequent exams. The BF Group ($N = 108$) did not receive the additional two questions on their survey due to the fact that they did not receive the elaborate feedback manipulation. Their responses were included in the original 18-question survey offered to all sections which assessed global levels of satisfaction with the course.

Discussion

Previous literature has supported the efficacy of providing feedback to improve performance (e.g., Alvero et al. 2001; Balcazar et al. 1985; Bangert-Drowns et al. 1991; Kulhavy 1977; Peeck et al. 1985; Vollmeyer and Rheinberg 2005) as well as PSI-type courses with respect to student performance (e.g., Austin 2000; Buskist et al. 1991; Caldwell et al. 1978; Kulik et al. 1979, 1980; Sherman 1992). The current investigation sought to answer the question as to what type of feedback (if any) would yield differential effects in terms of improved student performance in a modified PSI course.

Results from the present study support and extend previous research that has suggested that the *type* of feedback delivered is related to subsequent student performance. For instance, Bangert-Drowns et al. (1991) conducted a meta-analysis which further examined different variables as they relate to feedback. Relevant to the current investigation was the type of feedback evaluated. The authors established that the *type* of feedback was strongly related to effect size. In particular, Bangert-Drowns et al. (1991) found that, "When feedback merely indicated that a response was correct or incorrect; it resulted in a lower effect than when the feedback in some way informed the learner of the correct answer" (p. 232).

The basic feedback and elaborate feedback manipulation implemented in this study closely parallels Bangert-Drowns et al.'s (1991) definitions of feedback that was corrective and explanatory, respectively. Results from the current investigation suggested positive effects in both the basic and elaborate feedback conditions for each group. That is, improvement was observed on subsequent attempts regardless of the feedback type received suggesting possible practice effects. However, significant differences were identified between the BF Group and the EF Group. Overall, there was a significant difference in performance and learning gain for individuals who received elaborate feedback relative to basic feedback.

In addition, an unexpected finding emerged with respect to the two types of feedback on the easier questions. Exposure to elaborate feedback resulted in statistically significant improved performance on these questions. It is assumed that subsequent performance on easier questions would be undifferentiated across the two types of feedback received. That is, if feedback's primary instructional significance is to correct errors (e.g., Kulhavy 1977), then it is further assumed that students would not spend additional time reviewing feedback on questions they had already answered correctly. Alternatively, it is possible that participants had answered these questions correctly but were not confident as to why. In this study, elaborate feedback was provided to the EF Group regardless of whether questions were answered correctly or not. Conversely, participants who received basic feedback were not provided with any additional information as to why a given answer was answered correctly. This may account for the difference found between the two types of feedback on the easier questions.

What may be of more interest is the significant difference that was found between elaborate and basic feedback on the most-difficult questions. Significant differences in performance were observed between the BF Group and the EF Group on the top quartile of the most-difficult questions. Thus, it is assumed that elaborate feedback benefited participants who answered these questions incorrectly and did not understand why. While there was a notable difference between the main effects of feedback type, it is important to acknowledge the influence interaction effects may have on main effects. That is, the effect of either basic or elaborate feedback may be differentially influenced by the relative level of question difficulty. Results of this study demonstrated a significant interaction between factors (i.e., feedback type and question difficulty) with elaborate feedback on the most difficult questions resulting in significantly improved performance relative to basic feedback.

Moreover, what may be of primary importance to most instructors and university administrators is the overall grade distribution of their students. When all other ancillary, academic scores had been factored out, the grade distribution of each of the two groups in the current investigation clearly favored those that received elaborate feedback. Although the BF Group received the lowest percentage of A's and the highest percentage of F's across each of the two groups, it should also be noted that the grade distribution for this group was negatively skewed with almost half (49%) of the participants receiving an A or a B. This approximates grade distributions in other PSI-type courses that are negatively skewed (Sherman 1992).

In general, the results of this study support the differential effectiveness of elaborate feedback relative to basic feedback on student performance. However, some limitations of this investigation should be addressed. A primary confound in the current investigation concerns the evidence of practice effects. That is, participants were exposed to identical, experimental questions two times and performance typically improved with subsequent exposure to most quiz items regardless of the type of feedback delivered. Thus, practice effects may have resulted by the mere exposure to test stimuli on multiple attempts. However, it should be mentioned that although identical questions were delivered in the same order for each of the groups, the order of the multiple-choice answer options were delivered randomly. Moreover, even though each group was exposed to identical questions the same number of times, differential results were obtained. Therefore, while improvement on subsequent attempts was observed across these groups, the elaborate feedback manipulation can account for the variability in relative improved performance.

While the present results demonstrated that exposure to elaborate feedback improved subsequent performance, no controls were implemented to ensure that participants actually reviewed the feedback component. However, results from the course survey suggested that a majority of participants in the EF Group did in fact contact the feedback manipulation in this study. In addition, the current study sought to investigate the differential effects of two types of *automated feedback* on student performance. The current structure of this educational system allows students to attend discussions that compliment their various schedules. No controls were included requiring participants to attend discussions offered by only one instructor. Although, all instructors followed a standardized teaching protocol, no controls were implemented to account for the idiosyncrasies of the instructors or the number of discussions participants could attend which may have differentially influenced performance.

To address these issues, future research could attempt to control for practice effects by selectively incorporating conceptually similar questions with similar levels of difficulty to assess learning gain. For example, future research could evaluate learning gain by providing questions on the second attempt that are topographically dissimilar (different phrasing) but having comparable answers. Another option would be to hold the question items constant (identical phrasing) while assessing learning gain on multiple-choice answer options that are topographically dissimilar but functionally equivalent (conceptually similar answers). In

addition, future researchers can program greater control by requiring participants to attend discussions offered by only one instructor as well as holding the number of discussions students can attend constant across groups.

It appears that providing some form of feedback is better than no feedback at all and students benefit more from receiving elaborate feedback relative to basic feedback. The primary components of elaborate feedback in this study included providing specific information regarding each question and more general conceptual information. It is possible though that providing too much information can have iatrogenic effects on student performance. However, a component analysis was not conducted to identify whether either *specific feedback* or *general feedback* were necessary or sufficient. Therefore, a component analysis of these aspects of the study would be an intriguing area of exploration for future research to identify the most effective form of elaborate feedback. Finally, future research could incorporate positive behavioral contingencies to ensure that students review the feedback component following each quiz and exam.

In conclusion, results of this study suggest that providing students with elaborate feedback can have a significant impact on performance in a large-enrollment course. Further, the current procedure introduced in this investigation enables an instructor to provide every student with immediate, individualized, and elaborate feedback regarding their performance which is typically not possible in large-enrollment courses. Furthermore, while feedback in general has been shown to be effective in improving performance, item analyses from the current investigation suggested that elaborate feedback may be more effective in terms of improved performance associated with difficult questions. With respect to a cost-benefit analysis, the initial response effort required to program elaborate feedback for questions in a large test bank is far outweighed by the benefit one's students realize in terms of learning.

According to Skinner (1968), effective instruction can be achieved without resorting to aversive control. We must arrange special contingencies to facilitate learning. As such, the role of instructor becomes that of a contingency manager or an educational engineer. In addition, it should be remembered that, "The student is always right" (Keller 1968, p. 88). That is, if our students are not learning, we need to evaluate our instructional practices and rearrange the appropriate contingencies so all meet the standards of higher education. As Skinner (1968) proposed, "The most effective techniques of instruction will be drawn only from the fullest possible understanding of human behavior, a goal toward which an experimental analysis slowly but steadily moves" (p. 226). An empirically based understanding of the conditions under which feedback is most effective and the subsequent application of this information may be one means to achieve this formidable end.

Acknowledgments We would like to thank Donald Karr, MA, BCBA for his assistance with the statistical analyses conducted in this study.

Appendix 1

(1 point) *Elaborate Feedback*

A psychologist asks 3-year-olds to explore a room filled with toys either with or without their mothers present. The time the children spend exploring is measured. Exploration time is thus an example of a(n) _____ variable, and the presence or absence of the mother is a(n) _____ variable.

Percent Value	Correct Response	Student Response		Answer Choices
0%			a	independent; independent
0%			b	dependent; dependent
0%		X	c.	independent; dependent *Specific Feedback:* **Incorrect. The time the children spend exploring is measured (the dependent variable), while the presence or absence of the mother is manipulated (the independent variable).**
100.0%	X		d	dependent; independent

General Feedback: **The dependent variable in an experiment is the variable that is believed to be dependent upon (affected by) another variable (the independent variable). In psychological experiments, it is usually some measure of behavior (p. 33). The independent variable in an experiment is the condition that the researcher varies in order to assess its effect upon some other variable (the dependent variable). In psychology, it is usually some condition of the environment or of the organism's physiology that is hypothesized to affect the individual's behavior. (p. 34)**

Appendix 2

(1 point) *Basic Feedback*

A psychologist asks 3-year-olds to explore a room filled with toys either with or without their mothers present. The time the children spend exploring is measured. Exploration time is thus an example of a(n) _____ variable, and the presence or absence of the mother is a(n) _____ variable.

Percent Value	Correct Response	Student Response		Answer Choices
0%			a	independent; independent
0%			b	dependent; dependent
0%		X	c.	independent; dependent
100.0%	X		d	dependent; independent

References

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985–1998). *Journal of Organizational Behavior Management, 21*, 3–29.

Arnett, P. P. (1985). Effects of feedback placement and completeness within Gagne's model for computer assisted instruction lesson development on concept and rule learning. *Dissertation Abstracts International, 46*, 2537A.

Austin, J. L. (2000). Behavioral approaches to college teaching. In J. Austin & J. E. Carr (Eds.), *Handbook of applied behavior analysis*. Reno, NV: Context.

Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7*, 65–89.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effects of feedback in test-like events. *Review of Educational Research, 61*, 213–238.

Braksick, L. W. (2000). *Unlock behavior, unleash profits*. New York: McGraw-Hill.

Bumgarner, K. M. (1984). Effects of informational feedback and social reinforcement on elementary students' achievement during CAI drill and practice on multiplication facts. *Dissertation Abstracts International, 45*, 1102A.

Buskist, W., Cush, D., & DeGrandpre, R. J. (1991). The life and times of PSI. *Journal of Behavioral Education, 1*, 215–234.

Caldwell, E. C., Bissonnettee, K., Klishis, M. J., Ripliey, M. M., Farudi, P. P., Hochestetter, G. T., et al. (1978). Mastery: The essential essential in PSI. *Teaching of Psychology, 5*, 59–65.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Orlando, FL: Holt, Reinhart & Winston.

Cross, L. H. (1995). Grading students. *Practical Assessment, Research & Evaluation, 4*, 1–6.

Daniels, A. C. (1994). *Bringing out the best in people*. New York: McGraw-Hill.

Eikenhout, N., & Austin, J. (2005). Using goals, feedback, reinforcement, and a performance matrix to improve customer service in a large department store. *Journal of Organizational Behavior Management, 24*, 27–62.

Fairbank, J. A., & Prue, D. M. (1982). Developing performance feedback systems. In L. W. Frederiksen (Ed.), *Handbook of organizational behavior management* (pp. 281–299). New York: Wiley.

Farragher, P., & Szabo, M. (1986). Learning environmental science from text aided by a diagnostic and prescriptive instructional strategy. *Journal of Research in Science Teaching, 23*, 557–569.

Heald, H. M. (1970). The effects of immediate knowledge of results and correlation of errors and test anxiety upon test performance. *Dissertation Abstracts International, 31*, 1621A.

Henrysson, S. (1971). Gathering, analyzing, and using data on test items. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 130–159). Washington, DC: American Council on Education.

Hickman, J. S., & Geller, E. S. (2003). Self-management to increase safe driving among short-haul truck drivers. *Journal of Organizational Behavior Management, 23*, 1–20.

Houmanfar, R., & Hayes, L. J. (1998). Effects of feedback on task completion, time distribution and time allocation of graduate students. *Journal of Organizational Behavior Management, 18*, 69–91.

Huberman, W. L., & O'Brien, R. M. (1999). Improving therapist and patient performance in chronic psychiatric group homes through goal-setting, feedback, and positive reinforcement. *Journal of Organizational Behavior Management, 19*, 13–36.

Keller, F. S. (1968). Good-bye teacher. *Journal of Applied Behavior Analysis, 1*, 79–89.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*, 211–232.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*, 79–97.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist, 34*, 307–318.

Kulik, C.-L. C., Kulik, J. A., & Cohen, P. A. (1980). Instructional technology and college teaching. *Teaching of Psychology, 7*, 199–205.

Pampino, R. N., Jr., MacDonald, J. E., Mullin, J. E., & Wilder, D. A. (2003). Weekly feedback vs. daily feedback: An application in retail. *Journal of Organizational Behavior Management, 23*, 21–43.

Peeck, J., Bosch, A. B., & van den Kreupeling, W. J. (1985). Effects of informative feedback in relation to retention of initial responses. *Contemporary Educational Psychology, 10*, 303–313.

Pressey, S. L. (1926). A simple device which gives tests and scores- and teaches. *School and Society, 23*, 373–376.

Prue, D. M., & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management, 3*, 1–16.

Rohn, D., Austin, J., & Lutrey, S. M. (2002). Using feedback and performance accountability to decrease cash register shortages. *Journal of Organizational Behavior Management, 22*, 33–46.

Rummler, G. A., & Brache, A. P. (1995). *Improving performance: Managing the white space on the organizational chart*. San Francisco: Jossey-Bass.

Sassenrath, J. M., & Gaverick, C. M. (1965). Effects of differential feedback from examinations on retention and transfer. *Journal of Educational Psychology, 56*, 259–263.

Sasson, J. R., & Austin, J. (2005). The effects of training, feedback, and participant involvement in behavioral safety observations on office ergonomic behavior. *Journal of Organizational Behavior Management, 24*, 1–30.

Sherman, J. G. (1992). Reflections on PSI: Good news and bad. *Journal of Applied Behavior Analysis, 25*, 59–64.

Sidman, M. (1960). *Tactics of scientific research. Evaluating experimental data in psychology*. Boston, MA: Authors Cooperative.

Skinner, B. F. (1938). *Behavior of organisms*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1968). *The technology of teaching*. Englewood Cliffs, NJ: Prentice-Hall.

Sulzer-Azaroff, B., & Mayer, G. R. (1991). *Behavior analysis for lasting change*. Fort Worth, TX: Harcout Brace College.

Thorndike, E. L. (1913). *Educational psychology, Vol. 1: The psychology of learning*. New York: Teachers College, Columbia University.

Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction, 15*, 589–602.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203–214.