International Journal of Teaching and Learning in Higher Education 2023, Volume 35, Number 1, 48-66
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Evaluation of Teacher, Self-Assessment, and Combined Feedback to Increase Students' Behavioral Observation Skills

Jie Zhang & Marcie Desrochers
State University of New York (SUNY) Brockport

Michael Fensken
University of Nevada Las Vegas

Feedback is an important component of online instruction. Few experiments have examined combined types of feedback (teacher feedback or student self-assessment) compared to extra practice with each on student learning. In this study, the experimental effectiveness of four methods of feedback delivery on students' acquisition of behavioral observation skills was evaluated—teacher feedback only (TT condition), student self-assessment only (SS condition), teacher-then-student self-assessment (TS condition), or student self-assessment-then-teacher (ST condition). The results, both for a random sample of individual learners as well as at an overall group level, suggest that the SS condition was superior to the TS condition. In the SS condition, not only did participants perform better but also their accuracy of self-assessment on strengths was positively correlated with their Behavioral Observation Post-Training Assignment 1 scores and total Post-Training Assignment scores. Across all conditions, participants perceived their feedback experience as positive. Further research is needed to determine whether additional training scenarios and variations in training methodology are necessary to promote students' learning.

Online instruction is a common mode of delivery for students learning college-level material. The National Center for Educational Statistics (NCES, 2021a) reports that among 19.6 million college students, 3.9 million completed at least one online course (19.7%), and 3.5 million learned exclusively online (17.6%). Due to the COVID-19 pandemic since March 2020, the number of online courses has rapidly increased globally, providing a potentially safer environment in which college students can continue their learning (UNESCO, 2020). Specifically, 84% of undergraduates reported taking some or all online courses (NCES, 2021b). Given the increasing demand for online learning, it is critical for instructors to identify effective methods to optimize students' acquisition of course material.

Student feedback is an integral part of most educational experiences. Feedback can be defined as "the means by which [the student] can at each stage of a course gauge how they are doing in terms of the knowledge, skills, and understanding that will determine their overall result in the course" (Scott, 2014, p. 56). There are a variety of methods for delivering feedback to students to facilitate their acquisition of course material.

Teacher Feedback

Teacher feedback delivered for students' answers is probably the most frequently delivered instructional consequence in education (Erdemir & Yeşilçınar, 2021). Teacher feedback has been defined as "all of the comments made by the teacher as a reaction to any activity or behaviour by the pupil" (Hargreaves, 2014, p. 295). Students prefer feedback delivered by their teacher rather than by peers, and want it to both contain information about what they are doing well and not so well (Scott, 2014). To be effective, such feedback must be timely, directive, and specific (Higgins et al., 2002). Elaborative teacher feedback (defined as the correct response plus presentation of the original instructional content and a verbal comment) is more effective for student learning rather than simply identifying the correct response or no feedback at all (Smith et al., 2019). Narciss (2013) suggests that the essential components of productive feedback include verification of the performance results as well as the elaboration of additional information, such as task requirements, concepts, mistakes, procedures, and/or cognitive processes.

In addition to elaborative teacher feedback, teaching presence or establishing the social and cognitive connections between students and the instructor is also recognized as important (McNeill et al., 2019). Teaching presence may motivate students to delve into the instructional material (Pan & Shao, 2020). Students report that a perceived high level of teaching presence may lead to learning gains, engagement, and positive feelings toward the course (Jung & Lee, 2018). One way to foster teaching presence in an online course is to provide frequent, direct, and timely teacher feedback to students (Wang et al., 2021).

Despite the benefits of teacher feedback in providing information to guide students' efforts and establish teaching presence, there are downsides to it as well. One problem is that students may not act on teacher feedback for a variety of reasons. Teacher feedback may be perceived negatively by students or those who are satisfied with their grades may lack the motivation to review teacher feedback (Crisp, 2007). Additionally, students may not respond to feedback if they lack a history of associating the feedback with improvement in future learning or grades in the course (McCune & Hounsell, 2005). There may also be a disconnect

between the teacher's feedback and students' receipt of it that decreases its effectiveness (Price et al., 2010)). Teachers may focus on errors and error correction whereas students may view red marks and teacher commentary punitively. To ameliorate these problems, instructions about feedback delivered by the teacher should communicate the constructive and useful function that knowledge of errors serves for the learner (Dweck & Yeager, 2019).

Another issue surrounding the provision of feedback involves the amount of time needed for a teacher to provide individually tailored feedback to a large class (60 or more students) or multiple courses (Price et al., 2010). The time-consuming nature of teacher feedback has led to some proposing use of self-assessment (Jamrus & Razali, 2019).

Student Self-Assessment

Student self-assessment refers to a variety of activities performed by the student on their own work to improve their learning and/or skills (Andrade, 2019). In its simplest form, self-assessment involves grading one's work by comparing it to a model or answer key also known as self-evaluation (Panadero et al., 2017). Another form of self-assessment involves students' comprehensive analysis of their strengths and weaknesses with qualitative judgments about their performance (Eva & Regehr, 2005). As a process, self-assessment may consist of the student: (a) comparing their work to the teacher's rubric; (b) seeking self-directed feedback (e.g., asking for feedback from teacher or peers); (c) engaging in self-evaluation; (d) performing self-reflection; and (e) setting goals for future iterations (Yan & Brown, 2017).

There are many potential advantages of a self-assessment approach, including that it is amenable to an online format and feasible with large class sizes. In a recent meta-analysis focusing on self-assessment, Andrade (2019) reported that the majority of studies included in the review showed a positive relationship between student self-assessment and achievement.

In addition to producing favorable academic outcomes, other significant skills may be fostered by students using self-assessment. During the self-assessment process students actively engage in problem-solving, goal setting, and controlling their own behavior and future direction rather than relying on others (Wehmeyer, 1995). Promoting students' self-determination and self-regulation of learning in this way may increase their awareness of the purpose and process of their own learning (metacognitive skills), and positively affect their motivation and engagement (Andrade, 2019).

Despite its potential benefits, there are limitations to students' self-assessment of their own learning when summative evaluation is at stake. A major issue is that students tend to award themselves a higher grade compared to teachers (Andrade, 2019), and this discrepancy may occur among those with poor grades or those motivated to obtain the highest grade possible (Tejeiro et al., 2012). As Falchikov and Boud's (1989) meta-analysis of student self-assessment found, the degree of concordance between students' and teacher's assessments may depend on the quality of the research (well-designed studies showing less discrepancy) and amount of students' experience with a subject matter (upper-level students' self-assessments showing more agreement with teacher's ratings). Unless it occurs for formative assessment, Seifert and Feliks (2018) suggest that the teacher be involved in the process of determining performance criteria and guiding students in their self-assessment to circumvent the possible inaccuracies that may occur. Another problem with self-assessment is that it may not be as effective as teacher feedback. In an in-class experiment, students' acquisition of a complex behavior observation task was better in the teacher feedback compared to that in a self-assessment condition (Desrochers et al., 2019). Chang et al. (2012) found that teachers scored more rigorously compared to peer- and self-assessment conditions with a learning portfolio task. In contrast, using a quasi-experimental design, Gibbs and Taylor (2016) found no difference in students' scores nor preferences between teacher versus self-assessment conditions in an online statistics course. These outcome inconsistencies across studies suggest that additional research is needed on this topic.

Combined Feedback

Administering both teacher and self-assessment feedback may synergistically be more effective than using either alone due to each compensating for weaknesses of the other. Combining teacher feedback with self-assessment would allow teacher modeling of evaluation behaviors which may foster student autonomy and metacognitive skills. With experience and teacher feedback, students may learn to appropriately self-grade. It may be that, as Wanner and Palmer (2018) recommend, formative self-assessment with teacher feedback is optimal for students' learning achievement.

Only a few studies have examined combined feedback on learners' performance. For instance, To (2021) implemented a student-centered approach involving a sequence of peer feedback (student-to-student qualitative comments), student self-assessment (an audio-recorded evaluative response to the peer feedback), and then teacher feedback (based on student performance and self-assessment). Surveys and semi-structured interviews with the coding of open-ended responses and content analysis of the feedback were used to describe 35 Chinese teacher postgraduates'

experiences. Although participants questioned its quality, peer feedback was perceived as effective, convenient, and yielding diverse perspectives. Self-assessment was seen to enhance the students' self-evaluation and reflection of their performance with a recognition that there may not be objectivity and accuracy involved. Participants believed that the teacher's feedback was more welcoming and personalized as a result of the prior self-assessment and peer feedback.

In another preliminary evaluation of the effectiveness of combined feedback, the microteaching skills of 48 preservice teachers who received teacher feedback, then peer feedback, and lastly, self-reflection were examined (Erdemir & Yesilcinar, 2021). Teacher feedback and then self-reflection was perceived to be most useful. Like To (2021), Erdemir and Yesilcinar (2021) used a descriptive research design which does not allow causal interpretation of the findings. Experimental research is needed to determine the relative effectiveness of combined teacher feedback and student self-assessment compared to either alone.

Providing teacher feedback and then student self-assessment may increase students' motivation to learn (Pan & Shao, 2020). Alternatively, self-assessment and then teacher feedback may establish students' autonomy and self-evaluation skills (To, 2021). Additional practice at the task furthered by extra teacher feedback or extra self-assessment may also benefit students' learning (Lipko-Speed et al., 2014). Research is needed to evaluate whether teacher feedback and self-assessment combined, teacher feedback only, or student self-assessment only is most effective.

Behavioral Observation

Feedback, in some form, is essential for students learning new skills such as behavioral observation. Behavioral observation is the mainstay of teachers' and psychologists' assessment tools and is used to measure levels of children's challenging behaviors (e.g., out-of-seat, talk-outs), analyze the possible environmental reasons for the behavior, and/or evaluate the effectiveness of treatment programs (Jiang et al., 2019). Behavioral observation is a complex skill involving defining behavior, identifying an applicable measurement system, conducting observations of videorecorded or live sessions, performing interobserver reliability (IOR), evaluating the adequacy of IORs, analyzing the data and interpreting a possible cause of the behavior (Hojnoski et al., 2020). With few exceptions (e.g., Desrochers et al., 2019), identifying the most effective feedback method to facilitate students' acquisition of behavior observation skills has seldom been studied.

Current Study

The purpose of the current study was to experimentally evaluate and compare the instructional effectiveness of four formative methods of feedback delivery on students' acquisition of behavioral observation skills. The conditions tested include teacher feedback only (TT), student self-assessment only (SS), teacher feedback-then-student self-assessment (TS), and student self-assessment-then-teacher feedback (ST). To evaluate the accuracy of participants' self-assessment and whether self-assessment was associated with summative assignment outcomes, additional analyses were conducted with a small sample of participants in the SS, ST, and TS conditions. Specifically, the degree of correlation of students' accuracy of self-assessment of strengths and weaknesses of learning with Behavioral Observation Post-Training performance was calculated. Participants' subjective evaluations of the effectiveness of their feedback condition were also collected and compared across conditions.

Method

Participants were randomly assigned to one of four conditions to experimentally test the effects of teacher feedback, student self-assessment, and combined feedback approaches. Following viewing the behavior observation tutorial, participants conducted behavioral observations of two video training assignments with condition-related feedback provided following each. Next, participants conducted behavioral observations of two post-training videos and their responses were later scored by a trained coder. After participants received their summative evaluation, a subjective evaluation of the exercise was performed by all participants. Figure 1 illustrates the procedure used in this study.

Participants and Settings

After Institutional Review Board (IRB) approval, three researchers conducted this study online via the course learning management system Blackboard® in courses that included the topic of behavioral observation at a four-year public university in the Northeastern United States. The study was conducted during six semesters from Spring 2019 to Summer 2020 in three courses—Introduction to Special Education, Assessments for Special Education, and Psychology Research Methods. Two special education courses were delivered 100% online and the psychology course was offered as a hybrid course with online and in-person components. A convenience sample of the students enrolled in these three courses was recruited with a total of 132 students, including special education (n = 107, 81.1%) and psychology (n = 25, 18.9%) undergraduate

and graduate students. The students voluntarily participated in this research study and informed consent was obtained from each participant. A bonus point was given if the student chose to participate and other bonus-earning opportunities were offered to the students who opted out of this study.

One-way Analysis of Variance (ANOVA) analyses indicate that there was no statistically significant difference between the four feedback conditions for the participants' age (p = ns), GPA (p = ns), or hours spent in the course per week (p = ns). Table 1 presents the frequency and percentage of the participants' demographic characteristics by condition.

Materials

There were four professionally produced training and post-training videos on behavioral observations used with the publisher's permission, ranging in duration from 1:13 to 1:54 minutes and portraying four students' (e.g., Eddie, Ray, Shane, and Tracy's, respectively) challenging behaviors (e.g., disruptive, non-compliance, or bullying behaviors) at elementary to high school settings (Liaupsin et al., 2000). The four videos on behavioral observations were released to the participants in order: Eddie – Training 1 Video, Ray – Training 2 Video, Shane – Post-Training 1 Video, and Tracy – Post-Training 2 Video. The online behavioral observation tutorial consisted of a PowerPoint® and iSpring® presentation, mounted in the course learning management system, to provide participants with the standardized background knowledge required to complete the assignments successfully. The content of the Behavioral Observation tutorial includes behavioral definition, behavioral observation and recording strategy, interobserver reliability (IOR), reasons for low IOR scores, and hypothesizing the function of someone's problem behavior.

The assessment-related materials included: (a) four 8-question Behavioral Observation Assignments, two for the training phase and another two for the post-training phase, (b) two answer keys, one for each of the two assignments in the training phase with the same answer keys used in all feedback conditions, (c) two student self-assessment assignments, one for each of the two training assignments, and (d) a 7-point Likert scale subjective evaluation survey (Desrochers et al., 2019). Eddie's and Ray's videos were assigned in the Behavioral Observation Training phase and Shane's and Tracy's videos were assigned to the Post-Training assessments. Appendix A shows an example of one of four behavior observation assignments and its answer key, which were used by the teacher to provide feedback in the teacher feedback condition and given to the students in the student self-assessment condition.

A self-assessment assignment sheet was given to the participant together with the answer key to the participants in the student self-assessment condition (see Appendix B). The self-assessment sheet consisted of the following questions: (1) Based on the results of the assignment, what did you do well? (2) What skills do you still need to develop? (3) What have you learned during the process of this assignment which may contribute to your professional growth? (4) Based on what you know and can do, how will you apply this self-assessment to guide your improvement in the specific area(s)? Table 5 includes the seven questions used in the subjective evaluation survey. The participants accessed all materials online via Blackboard®, a course management system. The researchers also utilized a written script that specified the instructions given by the researcher to the participants and listed what procedural steps would occur in each condition as the study progressed.

Procedure

Research Design

A four-group posttest-only randomized experiment was used to evaluate the effect of type of feedback on participants' acquisition of behavioral observation skills. No pre-test was administered to reduce testing effects. Participants were randomly assigned to one of four feedback conditions: (a) self-assessment-self-assessment (SS) condition (n = 33, 25.2%), (b) self-assessment-teacher feedback (ST) condition (n = 33, 25.2%), (c) teacher feedback-self-assessment (TS) condition (n = 34, 26%), or (d) teacher feedback-teacher feedback (TT) condition (n = 31, 23.7%). All participants were informed that the content and assigned activities were part of their regular course content. Prior to training, participants were asked to complete an online Behavioral Observation tutorial that included examples of behavior definitions and recordings, as well as opportunities to practice observing and recording children's behavior, with model answers provided.

Training Phase

The assignments were completed by participants individually online within the given timeframe of one week identified by Researchers 1 and 2, the course instructors, for the course grade. The automatic adaptive release function of Blackboard® was used to ensure the participants in the four groups completed the correct sequence of activities. The participants were asked to first review Training Video 1 (e.g., Eddie's video) and complete the Behavioral Observation Training Assignment 1. Then, within 48 hours of the participant completing the assignment, teacher feedback or student self-assessment with the answer key was provided based

Figure 1
Flowchart of the Procedure of the Study

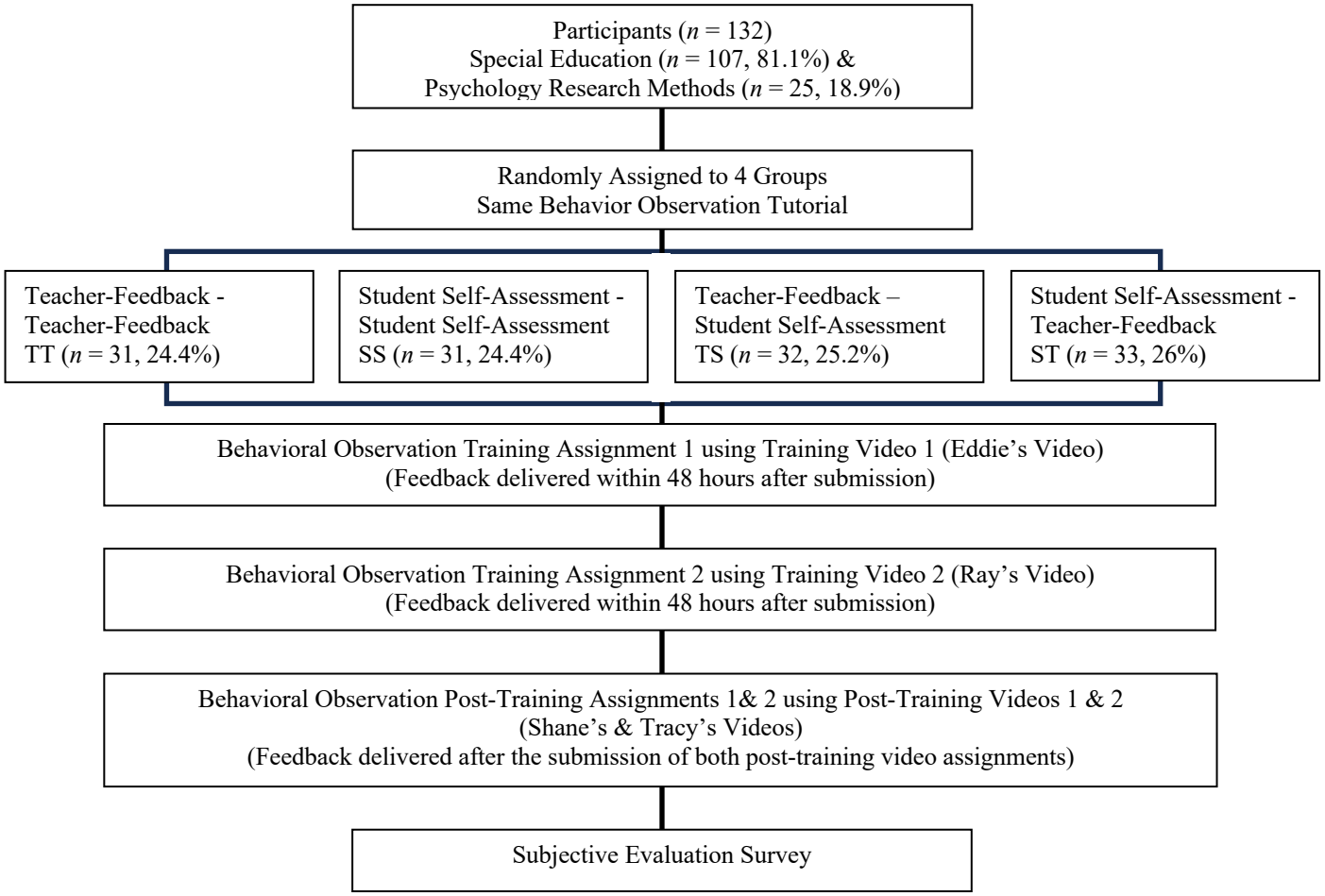


Table 1
Participants' Demographic Characteristics by Feedback Condition (SS, ST, TS, & TT

Aspect		Self-Self (SS)	Self-Teacher (ST)	Teacher-Self (TS)	Teacher-Teacher (TT)	Sub-Total
Discipline	Special Education	27	27	27	26	107
	Psychology	6	7	6	6	25
Gender	Female	21	23	20	28	92
	Male	10	10	13	4	37
	Gender-Fluid	1	1	0	0	2
	Not Specified	1	0	0	0	1
Mean Age (SD)		25.3	23.9	24.7	25.1	
Ethnicity	White	29	33	2	27	117
	Black/African-American	0	0	1	3	4
	Latino/Hispanic	0	0	2	1	3
	Asian	2	0	0	0	2
	Other	2	1	2	1	0
Year in College	Freshman	4	0	0	0	4
	Sophomore	4	2	5	5	16
	Junior	6	11	9	10	36
	Senior	4	7	7	7	25
	Graduate Student	15	13	12	9	49
General Grade Point Average (GPA)	2.0 - 2.5	2	1	1	3	7
	2.6 - 3.0	3	1	7	7	18
	3.1 - 3.5	4	17	13	9	43
	3.6 - 4.0	22	14	10	10	56
	NA	2	0	2	2	6
Weekly Hours Spent Studying	< 2	1	3	4	4	12
	3 - 5	15	19	17	15	66
	6 - 8	14	9	8	9	40
	> 8	3	3	4	4	14

on the condition of T or S to which participants were randomly assigned before accessing Training Video 2. After participants completed Training Video 2 (e.g., Ray's video) and its Behavioral Observation Training Assignment 2, within 48 hours the researcher provided feedback according to the condition to which participants were randomly assigned.

In the *teacher feedback* condition, the researchers provided teacher feedback to their own students using the same answer key and the scripted procedure online via Blackboard®. The written feedback consisted of the teacher marking a check mark (√+) next to the participant's correct answers to signify that the participant's answer is correct and adding brief comments (directives and explanations) next to incorrect answers on the participant's answer sheet (e.g., feedback for Q2: Frequency of the disruptive behavior: how many times Eddie showed the disruptive behavior described; feedback for Q8: Attention positive reinforcement, since at the beginning, behavior happened yet it was not related to any tasks. Eddie gets a reaction from others—both students and teacher when he engages in the disruptive behavior.).

In the *student self-assessment* condition, following submission of the Behavioral Observation Training Assignment 1, the participant: (a) received the assignment answer key, which is the same answer key as that used in the teacher feedback condition, to review and score their answers, and (b) was given the self-assessment sheet to reflect on their learning, strengths and areas for improvement. The participant then submitted their self-assessment before accessing the second video scenario. When reviewing the answer key, the participants self-assessed their performance by answering four questions as part of the assignment: (1) Based on the results of the assignment, what did you do well? (2) What skills do you still need to develop? (3) What have you learned during the process of this assignment which may contribute to your professional growth? (4) Based on what you know and can do, how will you apply this self-assessment to guide your improvement in the specific area(s)? The participants assigned in the self-assessment condition followed the same procedure when completing the Behavioral Observation Training Assignment 2.

Post-Training Assessment Phase

Following the training phase, all participants had one week to complete two behavioral observation post-training assignments, answering the same eight questions concerning behavioral observation as those in the training assignments. Participants watched and completed these behavioral observations for video 3 and video 4. Participants were also asked to fill out an eight-question subjective evaluation survey concerning their learning experiences.

Subjective Evaluation Data Analysis

Thematic analysis was used to identify, analyze, and document themes and patterns in the participants' answers to the subjective evaluation survey (Braun & Clarke, 2006; Saldaña, 2009). Researchers followed the thematic analysis procedure involving: (1) familiarizing themselves with the qualitative data by reading and re-reading the participants' answers, (2) generating initial codes by assigning preliminary codes to the data to describe the content, (3) searching for patterns and themes in codes across different students' answers, (4) reviewing themes, (5) refining codes, name themes, and created categories through ongoing analysis, and (6) reporting the results (Braun & Clarke, 2006). Following the initial coding, researchers met and reviewed the analysis and the themes. Researcher 1 checked for trustworthiness to strengthen the integrity of the findings (DeCarlo, 2019).

Coding Procedure

A trained researcher, "blind" to the participant's condition, scored each participant's answers for the assessments. The participant's answer to each question was compared to the answer key and given a score based on incorrect, partially correct, or fully correct for a total of five possible points for each assignment (Desrochers et al., 2019). Each assignment had a total of eight questions, each scored proportionately based on the number of correct descriptions given compared to the answer key. For example, students were asked to create an ABC chart of the scene unfolding in the video. There were a total of 24 unique antecedents (A), the child's behaviors (B), and consequences (C) for a particular video. If the participant answered more than 18 of the possible A, B, or C's (75%) correctly as compared to the answer key, they received one point, between 12 and 17 (75% and 49%), they received 0.75 points. For a question where participants wrote a short division problem to obtain the IOR, they were awarded a maximum of 0.5 points for a correct answer.

Subset Analysis

To better understand the accuracy of the students' self-assessment, four data sets from participants in the student self-assessment condition who were randomly selected and scored by the researchers, or 25% of the participants, were further analyzed. Data sets from eight participants' Training Assignment 1 and their Training Assignment 2 in the SS condition, from eight participants' Training Assignment 1 in the ST condition,

and those from another eight participants' Training Assignment 2 in the TS condition were used for analysis. The students' self-assessed strengths (i.e., "What did you do well?") and areas for improvement (i.e., "What skills do you still need to develop?") were compared to the individual scores they received on the assignment to determine the accuracy of students' self-assessments. If a student received one or 0.75 points from the researchers regarding a question they self-assessed as "strengths," it indicated that the student's self-assessment agreed with the researchers' independent grading of their answers, so they accurately self-assessed their strengths. Similarly, if a student received 0 or 0.25 points for a question they self-assessed as an "area for improvement," it indicated that the student accurately self-assessed the area for improvement. If a student received one or 0.75 points for a question that they self-assessed as an "area for improvement," or if they received 0 or 0.25 points for a question that they self-assessed as a "strength," it indicated that the student did not self-assess the question accurately. The percentage of the agreement between students' self-assessment and their performance scores, or the accuracy of the students' self-assessment on their strengths and areas for improvement were calculated across conditions (SS, ST, & TS) during the training and post-training phases. Pearson correlation coefficients were also used to measure the association between self-assessed strengths and Post-Training Assignment scores in the SS Condition.

Procedural Reliability

Procedural reliability was conducted to ensure the feedback procedure was correctly delivered. A random selection of 25% of the teacher feedback or the students' self-assessments across four conditions during the training phases (n = 32) was analyzed (eight students per condition). Three questions were asked when checking the procedural reliability: (1) Did the teacher deliver the answer key to the students in the self-assessment condition? (2) Did the students write viable answers to the four questions asked in the self-assessment condition? (3) Did the teacher provide feedback, which reflected correct answers, to the students assigned in the teacher feedback condition? The result of the analyses indicates 100% accuracy in the implementation of teacher feedback and students' responses to self-assessment questions.

Interobserver Reliability

Interobserver reliability (IOR) procedures were performed to determine accuracy in our coding of the post-training assessment data. Researcher 3 independently scored a randomly selected 25% of all participants' assessments. The IOR score was calculated based on one rater's smaller total score divided by another rater's larger score multiplied by 100. The average IOR score for Behavioral Observation Post-Training Assignment 1 was 93% ranging from 75-100% while that for Post-Training Assignment 2 was 90.26% ranging from 67-100%.

Results

Multivariate Analysis of Variance (MANOVA) analyses results showed a significant difference in participants' Behavioral Observation Post-Training Assignment 1 scores across feedback conditions (F (3, 126) = 2.77, p = 0.04), as well as a significant difference in participants' total scores for the combined Post-Training Assignments 1 and 2 across feedback conditions (F (3, 126) = 3.02, p = 0.03). There was no significant difference in participants' Post-Training Assignment 2 scores across feedback conditions. Table 2 reports Tukey's post-hoc test results. The mean score of the participants' Post-Training Assignment 1 scores in the TS condition (M = 3.53, SD = 0.92, SE = 0.13) was significantly lower than that in the SS condition (M = 4.04, SD = 0.54, SE = 0.13). In addition, the mean score of the participants' total scores for the combined Post-Training Assignments 1 and 2 total score in the TS condition (M = 7.19, SD = 1.54, SE = 0.22) was significantly lower than that in the SS condition (M = 8.07, SD = 0.97, SE = 0.22). Table 3 reports descriptive statistics, including the mean, standard deviation, and standard error of the Behavioral Observation Post-Training Assignment 1, Assignment 2, and combined Assignments 1 and 2 total scores by feedback condition.

Results of the Students' Self-Assessments Work Samples

Among the randomly selected 25% of participants across four conditions (n = 32), six students in the SS condition improved their performance in the behavioral observation post-training phases, followed by five students in the ST condition, four students in the TT condition, and two students in the TS condition. There was no significant difference found in the ANOVA tests across four conditions among the 25% selected samples (p = ns).

Subset Analysis

The results indicate that the percentage of the accuracy of the students' self-assessments on areas for improvement for Behavioral Observation Training Assignment 1 in the SS condition was the highest (56.25%), while the percentage of the accuracy of the students' self-assessments on strengths for Training

Assignment 1 in the ST condition was the lowest (0%). The percentage accuracy of the participants' self-assessments on areas for improvement for Training Assignment 1 in the ST condition was also relatively low (18.75%). The limited number of participants whose data were analyzed may be a factor (n = 8, or 25% of all participants, per condition). The relatively low accuracy of self-assessment among the randomly selected eight participants in the ST condition may have been insufficient to be generalized or lead to definitive conclusions. Table 4 indicates the percentage of the accuracy of participants' self-assessments on strengths and areas for improvement across conditions, which was calculated by the number of questions correctly self-assessed on strengths or areas for improvement being divided by the total number of self-assessed questions, and multiplied by 100%.

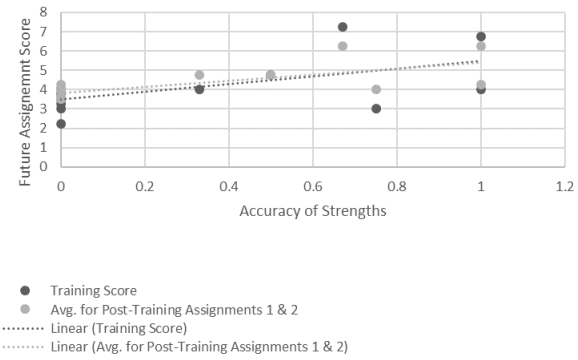
When analyzing the data in each condition, participants' percentage of accuracy of self-assessment on strengths regarding Behavioral Observation Training Assignment 1 was positively correlated with their performance in the Behavioral Observation Post-Training Assignment 1 in the SS condition. In the SS condition, the more accurately participants self-assessed on their strengths ($M = 0.26$, $SD = 0.4$), the higher score they received in the Post-Training Assignment 1 ($r(6) = 0.82$, $p = .012$, $M = 4.31$, $SD = 0.95$). Similarly, in the SS condition, participants' percentage of accuracy of self-assessment on strengths regarding Training Assignment 1 was positively correlated with their total score for Post-Training Assignments 1 and 2. The more accurately participants self-assessed on their strengths in the SS condition, the higher total score they received in the Post-Training Assignments 1 and 2 ($r(6) = 0.78$, $p = 0.023$, $M = 8.5$, $SD = 1.81$). Figure 2 shows the scatter plots of the accuracy of the students' self-assessment on strengths versus their performance in the training phases, and the average of the post-training performance for participants assigned in the SS-Training Assignment 1 and SS-Training Assignment 2 conditions.

Participants' Subjective Evaluations

Among 132 participants, 127 returned the subjective evaluation survey (96.2%), with 24.4% ($n = 31$) in the SS condition, 26% ($n = 33$) in the ST condition, 25.2% ($n = 32$) in the TS condition, and 24.4% ($n = 31$) in the TT condition. Multivariate Analysis of Variance (MANOVA) analyses indicate that there was no statistically significant difference in participants' ratings of the subjective evaluation questions between the four feedback conditions (p = ns). Table 5 shows the means and standard deviations of the participants' ratings of the subjective evaluation questions across conditions.

Fifty-four participants wrote additional comments on the subjective evaluation survey ($n = 54$, 42.2%). The following themes were identified: (1) practical experience ($n = 28$, 51.9%), (2) informative content ($n = 17$, 31.5%), (3) positive comments on feedback ($n = 8$, 14.8%), and (4) difficult questions ($n = 2$, 3.7%). In addition, 11 participants provided suggestions (20.4%). When compared across the feedback conditions, 26 participants in the ST condition provided comments (48.1%), followed by 18 in TS (33.3%), and then 16 in SS (29.6%). Six students in the TT condition provided additional comments (11.1%). Table 6 illustrates the themes and descriptive statistics of the participants' comments across conditions. Chi-square analyses indicate that there was no statistically significant difference in participants' coded comments for the subjective evaluation survey questions between the four feedback conditions (p = ns).

Figure 2 *Scatter Plot of the Linear Relationship Between the Accuracy of Students' Self-Assessed Strengths vs. Training Performance and Average of Post-Training Performance in the SS-Training Assignment 1 and SS-Training Assignment 2 Conditions*



The participants affirmed the practical and hands-on features of the project. One participant shared: "I found the behavioral assessment to be very beneficial to me and my professional career.... This has greatly influenced my teaching and allowed me to better help my students." Participants liked the informative content: "I am enjoying learning through the use of the videos and informational slides/materials and feel very well prepared...." The category of "positive comments on feedback" contained comments on timely, helpful, and detailed feedback. An example of one participant's comment is "I like how we were given feedback as to what we did wrong ahead of the next assignment so we knew what to look for and what mistakes not to make again." There was no difference in the content of the comments favoring the S or T conditions in the combined conditions (ST or TS).

In terms of the difficulty theme, two participants in the TS condition shared, "I found it somewhat difficult to truly know what measurement system.… I definitely

Table 2
Tukey's Post Hoc of Participants' Scores by Feedback Condition

Question	(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Behavioral Observation Post-Training Assignment 1 Score	TS	SS	-0.51*	.18	.03	-0.98	-0.04
Behavioral Observation Post-Training Assignments 1 & 2 Total Scores	TS	SS	-0.88*	.31	.03	-1.68	-0.07

Note. * indicates there was a significant difference at the 0.05 level.

Table 3
Descriptive Statistics of Behavioral Observation Post-Training Assignment Scores by Feedback Condition

Question	Condition	Mean	Standard Deviation	Standard Error
Behavioral Observation Post-Training Assignment 1 Score	SS	4.04	0.54	0.09
	ST	3.83	0.76	0.13
	TT	3.77	0.64	0.12
	TS	3.53	0.92	0.16
Behavioral Observation Post-Training Assignment 2 Score	SS	4.03	0.55	0.10
	ST	3.88	0.98	0.17
	TT	3.78	0.65	0.12
	TS	3.66	0.69	0.12
Behavioral Observation Post-Training Combined Assignments 1 & 2 Total Scores	SS	8.07	0.97	0.17
	ST	7.71	1.55	0.27
	TT	7.56	1.15	0.21
	TS	7.19	1.54	0.26

Table 4
Percentage of the Agreement Between Students' Self-Assessments and Performance Scores (or the Accuracy on Strengths and Accuracy on Areas for Improvement) Across Conditions (SS, ST, & TS)

Condition	Accuracy on Strengths (%)	Accuracy on Areas for Improvement (%)
SS - Behavioral Observation Training Assignment 1	26	56.25
ST - Behavioral Observation Training Assignment 1	0	18.75
SS - Behavioral Observation Training Assignment 2	30.95	18.75
TS - Behavioral Observation Training Assignment 2	41.63	16.63

Table 5
Participants' Ratings of the Subjective Questions by Feedback Condition (SS, ST, TS, & TT)

Question	SS Mean (SD)	ST Mean (SD)	TS Mean (SD)	TT Mean (SD)
Q1. How helpful was the instructional feedback given during the behavioral observation class? (1 = not at all to 7 = extremely helpful)	6.09 (1.03)	6.39 (0.77)	5.95 (1.13)	6.22 (1.18)
Q2. How much do you think you learned about behavioral observation methodology due to the assignment given during class? (1 = not much at all to 7 = a large amount)	6.19 (0.82)	6.32 (0.76)	6.19 (0.94)	6.06 (0.84)
Q3. How easy/difficult was the assignment for the behavioral observation class? (1 = very difficult to 7 = very easy)	4.47 (1.27)	4.21 (1.32)	4.50 (1.19)	4.42 (1.15)
Q4. Would you recommend the behavioral observation class to your friends and classmates? (1 = definitely would not to 7 = definitely would)	5.94 (1.29)	5.77 (1.37)	6.12 (0.91)	6.00 (1.21)
Q5. As a result of the behavioral observation class, how prepared were you for the post-assessment? (1 = not at all prepared to 7 = extremely prepared)	5.47 (1.19)	5.59 (1.28)	5.59 (1.05)	5.58 (0.92)
Q6. I usually don't do very well on tests and exams, even if I study. (1 = completely agree to 7 = completely disagree)	4.97 (1.62)	4.95 (1.76)	4.00 (1.78)	4.42 (1.65)
Q7. I enjoyed participating in the behavioral observation class. (1 = completely agree to 7 = completely disagree)	3.03 (1.94)	2.56 (1.82)	3.24 (2.10)	3.10 (1.99)

Table 6
Themes, Frequency of Occurrence, and Percent of the Participants' Comments Across Conditions (SS, ST, TS, & TT)

Themes	SS (*n*=16)	ST (*n*=26)	TS (*n*=18	TT (*n*=6)
Practical feature (n=28)	9	12	5	2
Informative content (n=17)	4	6	4	3
Positive comments on feedback (n=8)	0	4	4	0
Difficult questions (n=2)	0	0	2	0
Suggestions (n=11)	3	4	3	1

need more practice at this to feel comfortable since there is no pause or rewind in a real classroom." Another difficult question was to *"construct a good behavioral definition of …'s disruptive behavior*. " "I thought the most difficult part was attempting to operationally define behaviors."

Lastly, participants provided suggestions for improvement of the project. Students commented on feedback, "I think more tailored feedback would have been helpful rather than just an answer key to decipher where I went wrong as a student." Newer video scenarios were suggested, "I recommend the use of newer videos because of how technology is changing how people interact." More observation practice was suggested, "I kind of wish there were more than four observations within the class." A student enrolled in the Spring 2020 semester when courses were unexpectedly switched to 100% online in the mid-semester due to COVID-19 shared their preference for in-person learning, "I just have a hard time learning online, feel like it would've been a lot different in person."

Discussion

Participants' scores in the SS condition were superior to that in the TS condition for Behavioral Observation Post-Training Assignments 1 and 2 combined, and for Assignment 1 alone. Participants' accuracy of student self-assessment on strengths during the Behavioral Observation Training Assignment 1 in the SS condition was highly positively correlated with the Post-Training Assignment 1 scores as well as with their total Post-Training Assignments 1 and 2 scores. Participants in all conditions rated the feedback as very helpful and wrote positive comments about their feedback experience and, in general, felt that the training helped them learn an important skill. Overall, these results suggest that during formative assessment teachers might need to refrain from providing students with feedback and, instead, allow students to self-assess their work to optimize their learning of a behavioral observation task.

The result that participants' scores in the SS condition surpassed that in the TS condition agrees with Andrade's (2019) finding of a positive relationship between student self-assessment and achievement. The self-assessment process in the SS condition may encourage students' self-determination and self-regulation of the learning process as suggested by the finding that participants' accuracy of self-assessment correlated with Behavioral Observation Post-Training Assignment scores (Wehmeyer, 1995).

Participants' higher scores in the SS condition may have been due to the consistency in the type of feedback delivered. Delivering the same type of feedback for both training scenarios may have created less disruption of participants' attention and orientation to the task compared to a change in feedback format across training scenarios (Nevah-Benjamin et al., 2019).

Participants' lower scores in the TS condition than that in the SS condition may be due to several factors. The contrast in procedures between the Teacher Feedback versus Student Self-Assessment conditions provided in this particular order may have contributed to participants' poorer performance compared to that in the SS condition. Participants were not informed in advance which feedback condition they were to receive and thus may not have expected that the self-assessment condition would be delivered after the Teacher condition. Building self-determination or autonomy and self-regulated learning when deviating from teacher feedback may require student preparation and training. In the future, instructions, a rationale, and examples should be presented in advance to prepare participants for the type of feedback that will be provided and teach them how to engage in self-assessment.

Also, it is possible that participants experienced a negative emotional response (e.g., resentment) due to the switch from Teacher Feedback to Student Self-Assessment conditions compared to consistently receiving either of those conditions for both assignments. Past research has found that students value teacher feedback (Scott, 2014). Nonetheless, in our study, participants' comments about their experiences across the different conditions were positive and no one condition, combined nor consistent, stood above the other. However, participants' subjective evaluations were sought at the end of the study, and may have differed if they were gathered immediately after receiving each feedback.

Our results disagree with Desrochers et al.'s (2019) outcome where higher participants' correct responses occurred in the Teacher Feedback condition compared to the Student Self-Assessment condition despite many procedural similarities (i.e., same training materials, teachers, and self-assessment feedback). The current study was conducted entirely online with a delay in delivery of feedback by, at most, 48 hours. In contrast, Desrochers et al.'s (2019) study was in-class and feedback was more immediately provided during the class in which students completed the assignment. It is possible that the effectiveness of teacher feedback or teaching presence is reduced in an online environment where textual information is delivered compared to a complex array of stimuli delivered by a live teacher in close physical proximity to the student (e.g., intonation, facial expression, gestures). Additional research is needed to evaluate if verbal feedback, delivered by a teacher who is in the same physical space as the student, is more effective than online asynchronous textual feedback.

We found no significant effects for participants' correct answers in the TT and ST feedback conditions compared to that in the other conditions. To control for confounds, the teacher delivered standardized feedback (viz. the answer from the answer key sheet). However, the effectiveness of other types of teacher feedback (e.g., rephrasing the question, asking for a student's rationale, usable, and personable) may be an important consideration for future research (Dawson et al., 2019; Jacobs et al., 2022).

Limitations

There are several limitations associated with the current study. For instance, it is unknown whether participants in the teacher conditions (i.e., TT, TS, ST) actually read their feedback. There was no check on participants' comprehension of the content in the teacher's emailed feedback. In contrast, participants in the Student Self-Assessment condition were required to answer reflection questions submitted to their instructor and so would be more likely to have carefully scrutinized the answer key to address the questions posed. A review of the self-assessment worksheets verified that participants completely answered all questions. Additionally, an analysis of 25% of participants' self-assessment worksheets in the Student Self-Assessment conditions (SS, ST, TS) suggests that they accurately evaluated their performance and that it was significantly and highly positively correlated with post-training scores. One method to facilitate students' attention to teacher feedback is to create a checklist of improvement items based on the teacher feedback for students to use when completing the next assignment. Another strategy to ensure treatment integrity is to add incentives or bonus points for students' improvements in answers compared to deficiencies noted in the previous teacher feedback (Callender et al., 2016).

Additionally, participants' learning could have been affected by the set order of presentation of the training videos. For instance, participants in the TS condition may have performed poorly due to the second scenario being more difficult than the first. In future research, counterbalancing the order of scenarios should be employed to eliminate this concern.

Another limitation is that the amount of training provided may have been insufficient to completely establish participants' skills. We only used two training cases and, given that participants' overall average scores ranged from 71% (TS) to 80% (SS), more training seems necessary to attain a mastery criterion of 90% or better.

Future Research

Based on an item analysis of participants' assignments, more instruction may be needed for certain questions including behavioral definitions, ABC observations, and identification of the reason for the individual's problem behavior depicted in the scenario. Additional training scenarios and feedback may be needed for students to master this complex behavioral observation task as a whole. One participant noted in the subjective evaluations that more practice scenarios would have been helpful. In Lew et al.'s (2010) research, students' repeated self-assessments across a four-week interval did not improve despite being provided with extensive feedback from peers and peer tutors. It is unknown whether similar results would occur for students with this behavioral observation task if additional practice with some variation in feedback configuration (e.g., STST, TSTS, SSST, TSSS, TTSS, or TTTS). Researchers need to explore the relative effectiveness of these possibilities.

In addition, researchers of this study used a criterion-referenced grading system to investigate the effectiveness of different feedback delivery methods on students' acquisition of behavioral observation skills. In this system, the researchers set the maximum points for each assignment (that is, 5 points) and a standard of performance (that is, answer key), subtracted points from the defined maximum of 5 points, and assigned grades based on the individual performance of each student (U.S. Department of Education, 2008). There are other grading systems, including norm-referenced grading systems and alternative grading systems, such as pass-fail systems and non-graded evaluations (U.S. Department of Education, 2008). Each grading system has its pros and cons and the impact of different grading systems on student learning could be further explored for future studies.

Conclusion

As Wiggins (2013) notes, "Learners need endless feedback more than they need endless teaching" (p. 1). This study was to investigate the experimental effectiveness of four methods of feedback delivery on students' acquisition of behavioral observation skills—teacher feedback only (TT condition), student self-assessment only (SS condition), teacher-then-student self-assessment (TS condition), or student self-assessment-then-teacher (ST condition). It provides preliminary evidence that consistent student self-assessment feedback (SS) is more effective than combined feedback (TS) when teaching behavioral observation skills to students enrolled in online classes. The results, both for a random sample of individual learners as well as at an overall group level, suggest that the SS condition was superior to the TS condition. In the SS condition, not only did participants perform better but also their accuracy of self-assessment on strengths was positively correlated with their Behavioral Observation

Post-Training Assignment 1 scores and total Araining Assignment scores. Across all conditions, participants perceived their feedback experience as positive.

Student self-assessment has the advantages of being able to be delivered to large classes and is much less time-consuming to provide than teacher feedback. A formative self-assessment approach may not only improve students' knowledge and skills on the instructional topic, but also foster autonomy and self-determination.

References

Andrade, H. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*, 1–13. https://doi.org/10.3389/feduc.2019.00087

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning*, *11*(2), 215–235. https://doi-org.brockport.idm.oclc.org/10.1007/s11409-015-9142-6

Chang, C. C., Tseng, K. H., & Lou, S. J. (2012). A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a web-based portfolio assessment environment for high school students. *Computers & Education, 58,* 303–320. https://doi:10.1016/j.compedu.2011.08.005

Crisp, B. (2007). Is it worth the effort? How feedback influences students' subsequent submission of assessable work. *Assessment & Evaluation in Higher Education, 32*(5), 571–581. https://doi:10.1080/02602930601116912

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education*, *44*(1), 25–36. https://doi.org/10.1080/02602938.2018.1467877

DeCarlo, M. (2019). *Scientific inquiry in social work.* Pressbooks. https://scientificinquiryinsocialwork.pressbooks.com/

Desrochers, M., N., Zhang, J., Caron, S., & Steinmiller, J. (2019). An experimental comparison of the effect of teacher versus self-evaluation/self-reflection feedback on college students' behavioral observation skills. *Journal of Behavioral Education, 28,* 258–274. https://doi.org/10.1007/s10864-018-09313-6

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science*, *14*(3), 481–496. https://doi.org/10.1177/1745691618804166

Erdemir, N., & Yeşilçınar, S. (2021). Reflective practices in micro teaching from the perspective of preservice teachers: Teacher feedback, peer feedback and self-reflection. *Reflective Practice, 22*(6), 766–781. https://doi-org.brockport.idm.oclc.org/10.1080/14623943.2021.1968818

Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine, 80*(10 Supplemental), S46–S54. https://doi:10.1097/00001888-200510001-00015

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395–430. https://doi.org/10.2307/1170205

Gibbs, J. C., & Taylor, J. D. (2016). Comparing student self-assessment to individualized instructor feedback. *Active Learning in Higher Education*, *17*(2), 111–123. https://doi.org/10.1177/1469787416637466

Hargreaves, E. (2014). The practice of promoting primary pupils' autonomy: Examples of teacher feedback. *Educational Research*, *56*(3), 295–309. https://doi.org/10.1080/00131881.2014.934554

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education, 27*(1), 53–64. https://doi:10.1080/03075070120099368

Hojnoski, R. L., Missall, K. N., & Wood, B. K. (2020). Measuring engagement in early education: Preliminary evidence for the behavioral observation of students in schools–early education. *Assessment for Effective Intervention*, *45*(4), 243–254. https://doi.org/10.1177/1534508418820125

Jacobs, J., Scornavacco, K., Harty, C., Suresh, A., Lai, V., & Sumner, T. (2022). Promoting rich discussions in mathematics classrooms: Using personalized, automated feedback to support reflection and instructional change. *Teaching & Teacher Education*, *112*. https://doi.org/10.1016/j.tate.2022.103631

Jamrus, M. H. M., & Razali, A. B. (2019). Using self-assessment as a tool for English language learning. *English Language Teaching*, *12*(11), 64–73. http://dx.doi.org/10.5539/elt.v12n11p64

Jiang, Y., Capriotti, M., Beaulieu, A., Rooney, M., McBurnett, K., & Pfiffner, L. J. (2019). Contribution of the behavioral observation of students in schools to ADHD assessment. *School Mental Health: A Multidisciplinary Research and*

Practice Journal, *11*(3), 464–475. https://doi.org/10.1007/s12310-019-09313-5

Jung, Y., & Lee, J. (2018). Learning engagement and persistence in massive open online courses (MOOCS). *Computers & Education, 122*, 9–22. https://doi.org/10.1016/j.compedu.2018.02.013

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education, 35,* 135–156. https://doi:10.1080/02602930802687737

Liaupsin, C. J., Scott, T. M., & Nelson, C. M. (2000). Functional behavioral assessment: An interactive training module. *User's manual & facilitator's guide* (2nd ed.*)*. ERIC Clearinghouse.

Lipko-Speed, A., Dunlosky, J., & Rawson, K. A. (2014). Does testing with feedback help grade-school children learn key concepts in science? *Journal of Applied Research in Memory and Cognition*, *3*(3), 171–176. https://doi.org/10.1016/j.jarmac.2014.04.002

McCune, V., & Hounsell, D. (2005). The development of students' ways of thinking and practising in three final-year biology courses. *Higher Education, 49*(3), 255–289. https://doi.org/10.1007/s10734-004-6666-0

McNeill, L., Rice, M., & Wright, V. (2019). A confirmatory factor analysis of a teaching presence instrument in an online computer applications course. *Online Journal of Distance Learning Administration, 22*(4), 1–16. https://www.westga.edu/~distance/ojdla/winter224/mcneillricewright224.html

Narciss, S. (2013). Designing and evaluating tutorial feedback strategies for digital learning environments on the basis of the interactive tutoring feedback model. *Digital Education Review, 23*, 7–26. https://files.eric.ed.gov/fulltext/EJ1013726.pdf

National Center for Education Statistics. (2021a). *Digest of education statistics 2020*, *table 311.15*. https://nces.ed.gov/fastfacts/display.asp?id=80

National Center for Education Statistics. (2021b). *2019–20 National postsecondary student aid study (NPSAS:20): First look at the impact of the coronavirus (COVID-19) pandemic on undergraduate student enrollment, housing, and finances (preliminary data).* https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021456

Naveh-Benjamin, M., & Brubaker, M. S. (2019). Are the effects of divided attention on memory encoding processes due to the disruption of deep-level elaborative processes? Evidence from cued- and free-recall tasks. *Journal of Memory and Language*, *106*, 108–117. https://doi.org/10.1016/j.jml.2019.02.007

Pan, X., & Shao, H. (2020). Teacher online feedback and learning motivation: Learning engagement as a mediator. *Social Behavior & Personality: An International Journal*, *48*(6), 1–10. https://doi.org/10.2224/sbp.9118

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review, 22,* 74–98. doi: 10.1016/j.edurev.2017.08.004

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, *35*(3), 277–289. https://doi.org/10.1080/02602930903541007

Saldaña, J. (2009). *The coding manual for qualitative researchers.* SAGE.

Scott, S. V. (2014). Practising what we preach: Towards a student-centered definition of feedback. *Teaching in Higher Education*, *19*(1), 49–57. https://doi-org./10.1080/13562517.2013.827639

Seifert, T., & Feliks, O. (2018). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education, 44*(2), 169–185. https://doi.org/10.1080/02602938.2018.1487023

Smith, A. C., Ralph, B. C. W., MacLeod, C. M., & Smilek, D. (2019). Test feedback and learning: Student preferences and perceived influence. *Scholarship of Teaching and Learning in Psychology*, *5*(4), 255–264. https://doi.org/10.1037/stl0000140 (Supplemental)

Tejeiro, R. A., Gomez-Vallecillo, J. L., Romero, A. F., Pelegrina, M., Wallace, A., & Emberley, E. (2012). Summative self-assessment in higher education: Implications of its counting towards the final mark. *Electronic Journal of Research in Educational Psychology*, *10*(2), 789–812. http://dx.doi.org/10.25115/ejrep.v10i27.1528

To, J. (2021). Using learner-centered feedback design to promote students' engagement with feedback. *Higher Education Research & Development*, 1–16. https://doi.org/10.1080/07294360.2021.1882403

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2020). *COVID-19 educational disruption and response*. https://en.unesco.org/covid19/educationresponse

U.S. Department of Education. (2008). *Structure of the U.S. education system: U.S. grading systems.* https://www2.ed.gov/about/offices/list/ous/international/usnei/us/grading.doc

Wanner, T., & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: The crucial factors of design, teacher participation and feedback. *Assessment & Evaluation in Higher Education*, *43*(7), 1032–1047. https://doi-

org.brockport.idm.oclc.org/10.1080/02602938.2018.1427698

Wang, Y., Stein, D., & Shen, S. (2021) Students' and teachers' perceived teaching presence in online courses. *Distance Education, 42(*3), 373–390. https://doi.org/10.1080/01587919.2021.1956304

Wehmeyer, M. L. (1995). *The ARC's self-determination scale (SDS): Procedural guidelines.* Beach Centre on Disability.

Wiggins, G. (2013). Putting students at the center. *ASCD Express, 8,* 1. http://www.ascd.org/ascd-express/vol8/811-toc.aspx

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education, 42*(8), 1247–1262. https://doi.org/10.1080/02602938.2016.1260091

JIE ZHANG, PhD, is a Professor of Special Education at State University of New York (SUNY) Brockport, where she teaches special education courses and supervises student teachers. Her research focuses on the efficacy of interventions for children with disabilities, services for children and families with diverse cultural and linguistic backgrounds, and Scholarship of Teaching and Learning (SoTL) in higher education. She has published 30 peer-reviewed journal articles at the national or international level and three book chapters. She received SUNY Chancellor's Award for Excellence in Teaching and Open SUNY Online Teaching Ambassador Award.

MARCIE DESROCHERS, PhD, is a board certified behavior analyst at the doctoral level (BCBA-D) and a New York State Licensed Behavior Analyst. She is an Associate Professor of Psychology at State University of New York (SUNY) Brockport where she teaches Applied Behavior Analysis, Developmental Disabilities, Research Methods and other courses. Marcie conducts research on best teaching strategies, assessment and treatment practices with individuals with developmental disabilities, and environmental sustainability behaviors.

MICHAEL FENSKEN, BS, is a graduate student in the clinical psychology doctoral program at the University of Nevada–Las Vegas. His clinical and research interests are anxiety disorders in youth and addressing school refusal behaviors. He has taught introductory psychology courses at the undergraduate level.

Appendix A

Example of one of Four Behavioral Observation Assignments and its Answer Key.

Training Video 1: Behavioral Observation Assignment on Eddie with Answer Key (AK)

1. Construct a good behavioral definition of Eddie's disruptive behavior. (Make sure your definition is objective, measurable, complete, and specific.)

AK - Any 1 of:

- any episode of talking out loud without teacher permission nor raising hand;
- interrupting (verbally or physical contact) one or more other students while they are working;
- name calling (saying words that other students rate negatively) of other student(s) (e.g., "Dork," "Idiots");
- out of seat without teacher permission when it is not the end of class.

AND

- if the teacher makes a verbal statement to the student to which the student responds within 5 seconds in an inappropriate manner to the teacher or other students (e.g., makes a grimace [twisted facial reaction], negative comment [e.g., "No, I won't do that"]. [required]
  One episode is differentiated from another if there is more than 3 seconds between each occurrence

2. Indicate the aspect of behavior to measure.

AK - Frequency of the disruptive behavior: how many times Eddie showed the disruptive behavior described.

3. Detail the measurement system you plan to use to record the behavior and state why you selected it.

AK - Continuous recording-Frequency recording method, since the behavior is discrete, at low rate, and within a short duration.

4. Observe and record the behavior independently. Conduct an ABC observation and analysis of the child's behavior.

AK - Eddie showed 11 times of disruptive behavior during the observation.

A	B	C	I	N	D
A peer came in, walked by Eddie's desk, and knocked down his folder.	Eddie: "Excuse me."	Peer looked back and said: "Excuse you."	x		
Peer looked back and said: "Excuse you."	Eddie raised up his voice: "I excuse you…"		x		
Teacher came in: "Let's get started by passing the homework."	Eddie: "What homework?"	Teacher: "The homework that's due every Monday."	x		
Teacher: "The homework that's due every Monday."	Eddie: "Since when?"	Teacher: "Eddie, since the beginning of the school year, we have homework that's due every Monday morning."	x		
Teacher: "Eddie, since the beginning of the school year, we have homework that's due every Monday morning."	Eddie raised his voice: "Not since I've been here."	Teacher did not respond, collecting homework from other students.	x		
Teacher did not respond, collecting homework from other students.	Eddie: "Call me a liar?"	Teacher: "I'm not calling you a liar, I'm just telling you what it is."	x		
Teacher: "I'm not calling you a liar, I'm just telling you what it is."	Eddie raised up his voice: "You've never told me." Stood up from his seat, hands laying out, "You've never told me. How am I supposed to do the HW if you've never told me?"	Teacher approached toward him: "Just sit down. I want you to get out of your books…"	x		

Teacher approached toward him: "Just sit down. I want you get out of your books…"	Eddie sat down.				x
Teacher: "Pair up. Make sure you answer questions from pp…"	Eddie looked up at the teacher. Opened his book and turned the pages. Looked up at his teacher again. Turned to his peer: "I'm not going to work with you, you idiots."		x		
	Eddie read the text, frowned, try to write down on his notebook but did not, re-read, flipped the pages, looked at two peers cross the isle, stood up, picked up folders on the floor and tried to look at their paper.	Male peer covered his book.		x	
Male peer covered his book.	Eddie: "What's your problem, punk?"	Male peer: "Do it yourself."	x		
Male peer: "Do it yourself."	Eddie: "You think you could make me do the work?		x		
	Eddie sat back to his seat, asking the teacher: "What do you look at?"		x		

5. Calculate an appropriate Interobserver Reliability (IOR) score. Assume the other observer recorded that Eddie's disruptive behavior occurred 11 times during the observation.
AK - IOR=smaller number/larger number X 100;
11/11 X 100 = 100%

6. Evaluate the adequacy of the IOR score based on your data.
AK - If the IOR >/= 80%, it is acceptable.

7. If the IOR score is inadequate then state why and redo.
AK - Poor definition; poor training, motivation; observer drift

8. State why Eddie's disruptive behavior is occurring (e.g., attention positive reinforcement, escape negative reinforcement, etc.) and explain why you think so.
AK - Attention positive reinforcement, since at the beginning, behavior happened yet it was not related to any tasks. Eddie gets a reaction from others—both students and teacher when he engages in disruptive behavior.

Appendix B

Student Self-Assessment/Self-Reflection Questions

Please review the answer keys, summarize your performance in a deliberate and thoughtful manner, and answer the following questions to self-reflect your performance in order to guide your future learning. Areas for future improvement need to be considered.

1. Based on the results of the assignment, what did you do well?
2. What skills do you still need to develop?
3. What have you learned during the process of this assignment which may contribute to your professional growth?
4. Based on what you know and can do, how will you apply this self-reflection to guide your improvement in the specific area(s)?