Students' Perceived Usefulness of Computerized Percentage-only vs. Descriptive Score Reports: Associations with Motivation and Grades

Lia M Daniels & Okan Bulut

Department of Educational Psychology, University of Alberta

Please direct all correspondence to Lia Daniels at lia.daniels@ualberta.ca. This research was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) Development Grant, Grant/Award Number: 430-2016- 00039 and a Faculty of Education Coutts Clarke Research Award to the first author.

Postprint of manuscript: **Daniels, L. M.** & Bulut, Ok. (2019). Students' Perceived Usefulness of Percentage-only vs. Descriptive Score Reports: Associations with Motivation and Grades. *Journal of Computer Assisted Learning, 36,* 199-208. DOI: 10.1111/jcal.12398

Abstract

In computer-based testing (CBT) environments instructors can provide students with feedback immediately. Commonly, instructors give students their percentage correct without additional descriptive feedback. Our objectives were (a) to compare stu- dents' perceived usefulness of a percentage-only score report vs. a descriptive feed- back report in a CBT environment and (b) to test relationships amongst perceived usefulness, motivation, and exam performance. Using a semester-long repeated mea- sures design embedded into three real examinations, we found that students per- ceived the descriptive feedback report as more useful than the percentage-only report. However, there were no relationships amongst the usefulness of the score report and students' motivation or exam scores. Instead, previous performance was the strongest positive predictor of future performance. We discuss the effortful work required to create descriptive feedback reports with their utility and suggest ways instructors may better support students in using descriptive feedback reports when they are implemented.

Students' Perceived Usefulness of Computerized Percentage-only vs. Descriptive Score Reports: Associations with Motivation and Grades

1.1 Introduction

Feedback in a summative testing situation is a tricky thing: Students say they want it, but do not automatically use it. Instructors dedicate hours to providing it, even if they are not sure what to say. Moreover, students and instructors both seem to balance the demands for quality feedback with speed of delivery. These considerations can be simultaneously facilitated and complicated by technological advances in terms of computer-based testing (CBT). For example, students suggest that the most helpful feedback is written and verbal (Blair, Curtis, Goodwin, & Shields, 2013), neither of which are easily amenable to CBT environments. However, Blair and colleagues also showed that students want their feedback quickly, and CBT environments can allow feedback to be delivered immediately to students. The objective of this research was (a) to examine how useful students perceived two types of immediate score reports in a CBT environment (percent-only report vs. descriptive feedback report) and (b) to examine relationships between this perceived usefulness, preference, and students' motivation and subsequent exam performance.

1.2 Theoretical Framework: Assessment

Hattie and Timperley (2007, p. 81) define assessment "as information provided by an agent (e.g., teacher, instructor, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" and go on to offer a model of feedback designed to enhance learning. Within this model, instructors are positioned to provide feedback with the purpose of helping students reduce the disparity between their current and desired level of understanding or performance. Indeed there is a healthy body of literature documenting a wide range of strategies

instructors employ to try and improve students' learning performance through feedback. For example, Isbell and Cote (2009) showed that low-performing social psychology students who received a personalized email expressing concern and providing course resources following their first exam performed better on their second exam than similar students who did not receive such an email. Wickline and Spektor (2011) showed that ungraded practices quizzes accompanied by correct answers resulted in improved students' performance, but the same was not true for graded quizzes or those that withheld the answers. And in her classic experiment, Butler (1988) showed that comments-only feedback was better for achievement and interest that comments accompanied by grades.

Both instructors and students themselves are responsible for using assessment information to make progress. According to Hattie and Timperley, instructors can support students by helping them make sense of their feedback and students can improve by increasing their efforts, adjusting their strategies, or disengaging if that is a valid option. Hattie and Timperley go on to suggest that feedback can be targeted at four levels, namely, task, process, self-regulation, and/or person, and that the target has specific implications for the student. These targets apply to immediate score reports in a computer-based testing environment, just as much as in traditional classrooms.

1.2.1 Feedback in computer-based testing

Computer-based testing (CBT) has become the hallmark of 21st century assessment (Sireci & Zenisky, 2006) and allows instructors to provide students with scores immediately after their examination. Most immediate feedback in CBT is at the task level because it conveys information to students on how well they performed the task at hand (Hattie & Timperley, 2007). The two most common examples of this type of computer-based feedback are raw scores (i.e.,

sum or weighted sum of items answered correctly) or percent-correct scores (i.e., percentage of items answered correctly), both of which can be calculated automatically and presented immediately upon completion of the exam. Research has shown that when students are satisfied with the percentage shown immediately following their exam, they experience an increase in relief, pride, and hope and a decrease of anxiety and shame (Author, 2017a; 2019). The experience when students are unsatisfied with their score, however, is more negative for students.

Although they are the easiest type of score report, percentage-only reports are not the only option. In particular, if a negative emotional profile occurs when students receive an unsatisfying percentage-only report, then instructors may want to provide students with more descriptive feedback immediately following their computer-based exam in the hopes of creating a more positive feedback experience. Hattie and Timperley (2007) would consider a descriptive feedback report to be a type of process level feedback because it can be used to extract information related to not only current performance but also how more general processes unfold. Descriptive score reports have become a legitimate option for instructors because of advances in score reporting; however, the process of initially creating and then maintaining such forms can still be quite laborious. To create a descriptive score report, at the minimum, the instructor will need to ensure that the exam is appropriately blueprinted to specific content domains and then create feedback tagged to each domain. This investment of instructors' effort and time may be appropriate if the descriptive report is more useful for students, however, this question has not been tested empirically. Additionally, because students have different goals in their course, motivational variables such as achievement goals (Elliot, 1999) may be related to students' perceptions of usefulness (Kluger & DeNisi, 1996).

1.3 Achievement Goal Theory

According to Kluger and DeNisi (1996), "self goals" play an important role in how students interpret feedback and achievement goals (Elliot, 1999) would be one of the most common ways to conceptualize students' goals. Specifically achievement goal theory describes four different goals, anchored by two concepts, that describe reasons students engage in certain behaviors. First, students tend to define competence either according to intrapersonal standards that focus on personal growth – mastery goals – or according to interpersonal standards that focus on performance relative to others – performance goals. It has been suggested theoretically (Dweck, 2008) and shown empirically (Bråten & Strømsø, 2004) that mastery goals are commonly associated with belief that one can improve with the investment of effort and appropriate strategies. In contrast, performance goals are associated with beliefs that investment of effort implies lack of ability. Second, students tend to be motivated towards success through approach tendencies or away from failure through avoidance tendencies. When crossed four types of achievement goals result: mastery-approach goals are generally defined as a desire to gain competence; mastery-avoidance goals are defined as a desire to avoid being incompetent; performance-approach goals are defined as a desire to demonstrate competence relative to others; and performance-avoidance goals are defined as desire to avoid appearing incompetent.

Some researchers have used achievement goals as framework to study how students make sense of their learning experiences, including test scores and feedback. For example, using a one semester longitudinal design, Senko and Harackiewicz (2005) showed that students had stronger endorsement of mastery-approach and performance-approach goals following high exam scores and stronger endorsement of performance-avoidance goals following low exam scores. In an experimental design involving three feedback conditions (score-only, negative feedback, positive

feedback), they further showed that students reduced their endorsement of mastery-approach goals most when given negative feedback compared to positive or score-only feedback. Type of feedback had no effect on performance-approach or performance-avoidance goals. More recently Pekrun and colleagues (2014) used an experimental design to show that anticipated self-referential feedback predicted mastery-approach goals whereas anticipated normative feedback predicted both performance-approach and –avoidance goals.

In both of these instances, the researchers (Pekrun et al., 2014; Senko & Harackiewicz, 2005) traded ecologically validity for experimental rigor; however, recent advances in CBT allow for an ecologically valid test of relationships between achievement goals, different types of immediate score reports, and students' actual grades. We hypothesize that students with mastery-approach goals may find descriptive feedback particularly useful because it may help them better monitor their growth and learning compared to a percentage-only report. Alternatively, students with performance-approach goals or any type of avoidance goal may find feedback in the form of a percentage-only report useful because it denotes their specific level of performance and can be compared to others. These important considerations can be tested experimentally in an ecologically valid context because of CBT. Thus, we pursued the following research questions in the current study:

1. Does students' perceived usefulness of and preference for receiving a percentage-only report differ from descriptive feedback report?
2. Is there a significant difference in students' preference for receiving a percentage-only descriptive feedback report and descriptive feedback report?
3. Do achievement goals relate to students' perceived usefulness of the percentage-only and descriptive feedback reports?

4. Does the perceived usefulness of different score reports predict students' actual score on the subsequent exam, after controlling for achievement goals and previous test score?

2.1 Method

We embedded a semester-long repeated measures design into a large-size undergraduate course in order to examine relationships among two different types of immediate score reports and students' achievement goals and objective exam performance. The research design was approved by the ethics review board of the institution and informed consent was implied by completing the questionnaire.

2.2 Procedure and Participants

In the first week of the semester all students in three sections of a blended-delivery required undergraduate course (N = 264) on Adolescent Development and Learning were invited to complete a pre-test questionnaire measuring their achievement goals and demographic information. Students accessed the online pre-test questionnaire through a link posted on their class management system, eClass by Moodle©. Of the total, 220 students (49 male, 165 female, 2 non-binary, age *M* = 24, range 18-54 years) accessed the questionnaire and 131 answered the achievement goal questions (49 male, 80 female, 2 non-binary, age *M* = 24, range 18-54 years). The sections used a common course design, with each course utilizing the same textbook, in-class activities, and examinations. In this way, the sections were highly similar, despite each section being taught by a different instructor.

Over the course of the semester students wrote three non-cumulative multiple-choice exams worth 15%, 20%, and 20% of their final grades, respectively. Students signed up to write the exams in a computer-based testing centre during a flexible three-day window of time. There were three equivalent forms of the examination with questions drawn from an instructor-made

test bank addressing content from the required textbook and lectures (Author, 2017b). The exams were blueprinted so that items were linked to specific topics covered in the course. The first exam contained 60 questions from the first three weeks of the course, the second exam contained 70 questions from the second three weeks of the course, and the final exam contained 74 items from the remaining five weeks of the course. At the completion of each exam, students had the option to access some type of immediate score report. After Exam 1 all students could access a percentage-only score report that showed their percentage correct (0-100%). After Exam 2, all students could access a descriptive score report (described below and shown in Figure 1). After the third exam students chose to see either a percentage-only report or a descriptive feedback report. To encourage participation in the exam setting, we offered a $5 coffee card to all participants who completed the five questions presented following the score report for all three exams resulting in exam-based data for $n = 214$ students (41 male, 139 female, 2 non-binary, age $M = 24$, range 18-54 years). Data from all portions of the study were combined using students' university identification numbers, which were deleted at the earlier time possible resulting in a completely anonymized data file, which can be accessed by contacting the researchers upon reasonable request.

The final sample consisted of 112 students who completed the pre-test items and all items presented after each of the three exam sessions, representing 42% of all students in the course. We tested for differences between the 112 students who completed all sessions including the pre-test and 117-128 students who completed one or more exam portions but not the pre-test. According to Bonferonni-corrected independent samples t-tests ($p < .007$), the groups did not differ in terms of age, perceived usefulness of either score report, or objective exam scores. Of the 112 participants with full information, the average age was 24 years (range 18-54 years), 41

participants identified as male, 69 female, and two as non-binary. For any analyses including gender, the non-binary students were removed.

2.3 Measures

2.3.1 Achievement goals

In the pre-test, we used Elliot and Murayama's (2008) Revised Achievement Goals Questionnaire (AGQ-R, 12 items) to measure students' achievement goals. The AGQ-R is one of the most commonly used measure of achievement goals in post-secondary settings and has strong evidence of reliability and validity (Hulleman, Schrager, Bodmann, & Harackiewicz, 2010). Three items assessed each of mastery-approach goals (e.g., My aim is to completely master the material presented in my class; mastery-avoidance goals (e.g., My aim is to avoid learning less than I possibly could; performance-approach goals (e.g., My goal is to perform better than the other students; and performance-avoidance goals (e.g., My aim is to avoid doing worse than other students. Participants indicated their strength of agreement with the items on a 1 = strongly disagree to 5 = strongly agree Likert scale. Higher scores on each subscale indicate stronger endorsement of that type of achievement goal; however, there are no normed cut-offs for interpretations of the scores.

2.3.2 Types of immediate score report

We created two types of immediate score reports for use in the computer-based testing environment. The first report was a *percentage-only report* that displayed the students' percentage correct (0-100%). The second report was a *descriptive score report* that included several visualizations representing students' performance on each domain on the examination (see Figure 1). In addition, students received written comments tailored to their performance on each domain (see Figure 2). For example, when students scored above 80% in a domain they

received positive written feedback at the end of the report conveying the message that "You did an excellent job applying the content from the chapters to the applications of teaching!". In contrast, when students scored less than 80% in a domain they received constructive comments that pointed to specific topics that could use review, stating for example "Your understanding of the influence of multiple systems in adolescents' lives could use some further review. You may want to review content related to siblings, fighting, attachment, parenting, and systems" (Figure 3). Only specific topics on which the student scored less than 80% were listed in their feedback.

2.3.3 Usefulness of immediate score report and preference

Immediately after seeing the score report, students responded to the five following questions: (1) This feedback is helpful (2) This feedback encourages me to put in more effort (3) This feedback will help me perform better on my next exam (4) This feedback will help me reach my goals (5) This is the type of feedback I want after an exam. Participants indicated their agreement with each item on a 1 = strongly disagree to 5 = strongly agree Likert scale. Based on an exploratory factor analysis (see Results), we summed all five items to create a single "usefulness" score with higher scores indicating greater perceived usefulness of the type of feedback. As an indicator of preference, students were allowed to choose which score report they wanted to receive after the last exam and this became a dichotomous variable.

2.4 Rationale for Analyses

As preliminary analyses we ran an exploratory factor analysis for usefulness items and then examined the descriptives and reliability for all self-report variables and the exam scores. To answer the four research questions, we used a combination of descriptive and inferential statistics. First, we used paired *t*-tests to compare students' perceived usefulness of the percentage-only report to the usefulness of the descriptive report. We expected students to

perceive the descriptive feedback report as more useful than the percentage only report. Second, we expected more students to choose the descriptive feedback report over the percentage-only report when given the choice after the last exam. Third, we ran correlations and examined the zero order associations between report usefulness and achievement goals. Finally, we used three separate regression analyses to examine the impact of perceived usefulness of feedback from the previous exam (Step 3) on subsequent exam score (dependent variable) while controlling for age, gender, achievement goals (Step 1), and previous exam score (Step 2). We expected current exam scores to be significantly and positively predicted by previous score, mastery-approach goals, performance-approach goals, and perceived usefulness of the immediate score report.

3.1 Results

3.2 Preliminary Analyses

The results of the exploratory factor analysis are in Table 1 and show that the usefulness items loaded onto a single factor following each of the three exams. The internal reliability measured by coefficient alpha was also adequate. The four achievement goal scales performed similar to existing research with adequate reliabilities and approach goals having higher scores than avoidance goals.

3.3 Perceived Usefulness and Preference

Students perceived the immediate descriptive feedback report provided after Exam 2 as significantly more useful (M = 20.48, SD = 3.31) than the percentage-only report provided after Exam 1 (M = 18.89, SD = 3.78), t (99) = -3.87, $p < .001$, CIs [-2.44, -.79]. As further evidence of students' preference for the descriptive report over the percentage-only report, 95.5% of students chose the descriptive report when they were given the choice after Exam 3. As an

indicator of practically how meaningful this distribution is, 107 students selected the feedback form compared to 5 who selected to see their percentage only.

3.4 Correlations and Regression Analyses

Correlations for all variables are presented in Table 2. All achievement goals were positively correlated except mastery-approach and performance-avoidance. The scores on each exam were positively correlated (r =.61 to .68) suggesting that previous performance will be predictive of future performance. Contrary to what we expected, there were no statistically significant associations between achievement goals and perceived usefulness of either the percentage-only report or the descriptive report. Achievement goals were also not significantly correlated to any of three exam scores.

For Exam 1, we had no indicator of previous achievement or any type of response to a previous immediate score report to include in the regression analysis and as such the overall explained variance was small. Thus, we entered age, gender, and the four types of achievement goals into a single step predicting students' score on Exam 1. The results are presented in Table 3 and show that age was positively associated with Exam 1 score and performance-avoidance goals negatively predicted Exam 1 score, although they did not explain a significant percent of the variance.

For Exams 2 and 3, after entering the same variables in Step 1, we included score on the previous exam in Step 2, and students' perceived usefulness of the score report at the conclusion of the previous exam in Step 3. For Exam 2, age and previous score on Exam 1 were positive significant predictors of score. There was no association between achievement goals or perceived usefulness of the percentage-only score report on students' scores on Exam 2. For Exam 3, although age and performance-approach goals were positive significant predictors of exam score

in Step 1, these effects were fully mediated by the inclusion of previous score on Exam 2 in the final step of the model. Additionally, gender emerged as a significant positive predictor of Exam 3 scores such that women obtained higher scores than men.

4.1 Discussion

Instructors invest substantial amounts of energy and time in providing students with feedback. However, Latham and Locke (1991) caution researchers and instructors to be aware that "few concepts in psychology have been written about more uncritically and incorrectly than that of feedback.... Actually, feedback is only information, that is, data, and as such has no necessary consequences at all" (p. 224). Our results partially support this statement and yet can also provide hope for instructors who feel strongly about providing students with timely feedback generally and descriptive feedback specifically – both of which are greatly facilitated by the rise of computer-based testing facilities. In this discussion, we weigh the effortful work required to create descriptive feedback reports with their utility for improving student performance and suggest ways instructors may better support students in using descriptive feedback reports when they are implemented.

4.2 Perceived Usefulness and Preference of Reports: What Students Like

Two of our results support that students' perceive the descriptive feedback report as more useful than the percentage-only report. First, the mean usefulness score was higher for the descriptive feedback report than the percentage-only report. Second, when given the chance to choose, students overwhelmingly picked the descriptive feedback report over the percentage-only report. This choice is particularly compelling because it was made following the final exam – a time at which descriptive feedback would arguably be least relevant because there were no future performance opportunities. While these results both point to utility of and preference for

the descriptive feedback report over the percentage-only arguably we may have expected the perceived usefulness ratings to have a larger difference given the overwhelming decision to choose the descriptive feedback report after Exam 3. Future research may want to counterbalance the presentation of the two reports on Exams 1 and 2 because it is possible that the percentage-only report received artificially high usefulness scores simply because it was provided immediately following the exam. If the percentage-only report was presented after the descriptive feedback report students immediacy and usefulness may be less conflated.

Although the preferred usefulness for the descriptive report over the percentage-only report was in line with our expectations, the relationships with achievement goals ran contrary to our expectations. We had expected mastery-approach goals to be positively associated with usefulness of the descriptive report and the remaining goals to be positively associated with the percentage-only report when in fact there were no statistically significant associations at all. One explanation for this may be that students simply all desire as much information as possible about their performance regardless of their goals for the course. Because the descriptive feedback report included percentage, from a basic cost-benefit analysis there was really no reason to *not* select the descriptive report over the percentage-only report. This represents an important limitation of the study and one that could be addressed by future research in which score is withheld from the descriptive report for some period of time. The decision to divorce score and description, however, must also be balanced by the desire to retain the ecological validity of the summative testing environment. In short, regardless of their achievement goals for the course, all students appeared to find the descriptive report more useful than the percentage-only score report following a CBT exam. From the perspective of what students seem to “like” or subjectively find

useful, we would recommend that instructors create descriptive score reports so long as the process is not prohibitively time consuming.

4.3 Predictors of Exam Scores: What Students Use

Even though students reported the descriptive report as more useful than the percentage-only report, we found no evidence that students' perceived usefulness of immediate score report improved their future performance. In other words, we infer that although students found the descriptive report more useful and selected it more often than the percentage-only report they did not actually use it - or at least did not use it in a way that resulted in the improvement of their subsequent exam score. This is a logical inference that we did not formally test in this study design. One limitation of the current research is that we do not know how long students viewed the descriptive feedback report or how they made sense of the information provided. As has been recommended by others (e.g., Burke, 2009; Price, Handley, Millar, & O'donovan, 2010), we encourage future researchers to examine how students interact with the descriptive score report that they preferred over the percentage-only report. This will provide important insights into the relationship between perceptions of usefulness and performance. To support this, instructors can refer to Nicol's (2007) seven principles of good feedback practice to support self-regulated learning specific to multiple-choice exams.

Moreover, for instructors who choose to dedicate their time and resources to creating descriptive feedback reports, we strongly encourage them to pair that effort with a strategy to support students in working through the report and making an action plan based on the feedback. Hattie and Timperley (2007) suggest that effective use of feedback can be guided by three questions: Where am I going? How am I going? Where to next? In a CBT environment instructors may be able to add these questions onto the descriptive feedback report as a type of

guided reflection. As described decades ago by Kulhavy (1977), this type of process can help feedback "take on the form of new instruction, rather than informing the student solely about correctness" (p. 212). Although this was beyond the scope of the current study, the implications for researchers and instructors are important.

Another possible explanation for a lack of association between score report usefulness and subsequent exam performance is that the examinations in the course were non-cumulative. We argued that the descriptive report could be considered feedback targeting the process; whereas the percentage-only report clearly targeted the task only (Hattie & Timperley, 2007). For example, if students scored particularly poorly on one domain they may decide that their current study skills are satisfactory but need to be better applied to all content areas. In contrast, if their performance was lacking in many domains they may decide to change their study strategies more generally. Thus, although the descriptive report could be used generally to guide future studying, students would not have an opportunity to demonstrate their knowledge on the specific topics again and thus designing ways for the descriptive score report to be tied more explicitly to process could be beneficial for impacting future grades.

Of all the possible predictors, achievement on the prior exam was the largest significant predictor of subsequent achievement, reinforcing a large body of literature (e.g., Richardson, Abraham, & Bond, 2012; Schneider & Preckel, 2017). This result suggests that the common finding that students who do well continue to do well and students who do poorly continue to do poorly was not disrupted by the implementation of immediate score reports – percentage-only or descriptive. This reinforces the need to create fair, reliable, and equitable exams to ensure that biases are minimized and true learning will be accurately captured for all students.

For Exam 1, in which we had no indicator of previous performance, age and performance-approach goals were significant positive predictors of score and performance-avoidance goals was a significant negative predictor. This replicates much of the existing achievement goal literature that shows performance goals more directly predict scores on exams in post-secondary settings than mastery goals (e.g., Bipp & van Dam 2014; Hulleman et al., 2010; Mouratidis, Michou, Demircioğlu, & Sayil, 2018) and highlights the adaptive nature of performance-approach goals for achievement scores in a normative college setting. It is possible that these effects did not emerge for later exams because students' achievement goals became further removed from exams 2 and 3. Although not consistent across all exams, it also appears that older students and women may have higher scores than younger students and men. These results are important when considering the ways in which students prepare for exams and prioritize their education (e.g., Buchmann & DiPrete, 2006; Richardson, 1994).

4.4 Limitations and Directions for Future Research

In addition to the limitations mentioned above, the results of the current research need to be considered in light of three limitations. First, although the ecological design of our study embedded directly into a real course with real examination is a strength, it imposed certain constraints as well including not being able to ask more questions about how students interacted with the descriptive score reports. This underscores why it is important for researchers to work closely with instructors to match research to instructional priorities. Second, because there was no incentive tied to the pre-test the sample was somewhat restricted by having fewer pre-test scores than we anticipated. Although the small incentive worked well for the exam portions of the study, researchers need to continue to wrestle with ways to maximize full participation with the rights of participants. Third, the pre-test to measure achievement goals was completed at the

start of the course and only once. It may be that the non-significant associations between achievement goals and exam scores might be because of the time distance between the variables. Alternatively, it is possible that students' achievement goals changed over time. Although we had hypothesized that achievement goals may be related to the type of feedback sought, it is equally possible that different types of feedback may influence the achievement goals that students endorse at a given time. Indeed, Pekrun and colleagues (2014) showed that whereas mastery-approach goals were triggered by anticipated self-referential feedback both performance-approach and performance-avoidance goals were triggered by anticipated normative feedback. As such future researchers should measure achievement goals around each exam specifically.

In conclusion, the results of this research clearly indicate that students perceive a descriptive feedback report as more useful than a percentage-only report and will overwhelmingly pick the descriptive report if given the choice following a computer-based test. However, provision of immediate feedback, regardless of its perceived usefulness did not impact students scores on subsequent exams. Instructors need not feel pressured to create these types of reports if the process is overly painstaking because the reports themselves did not improve students' exam scores. Thus, our recommendation is that if instructors choose to dedicate the time and resources to creating descriptive feedback reports for use in CBT environments, they also dedicate the time and resources to ensuring students use the reports to plan for subsequent examinations.

References

Bipp, T., & van Dam, K. (2014). Extending hierarchical achievement motivation models: The role of motivational needs for achievement goals and academic performance. *Personality and Individual Differences*, *64*, 157-162.

Blair, A., Curtis, S., Goodwin, M., & Shields, S. (2013). What feedback do students want?. *Politics*, *33*(1), 66-79.

Burke, D. (2009). Strategies for using feedback students bring to higher education. *Assessment & Evaluation in Higher Education*, *34*(1), 41-50.

Buchmann, C., & DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. *American sociological review*, *71*(4), 515-541.

Bulut, O., Cutumisu, M., Aquilina, A. M., & Singh, D. (2019). Effects of digi- tal score reporting and feedback on students' learning in higher edu- cation. Frontiers in Education, 4(65), 1–16.

Bråten, I., & Strømsø, H. I. (2004). Epistemological beliefs and implicit theories of intelligence as predictors of achievement goals. *Contemporary Educational Psychology*, *29*(4), 371-388.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*(1), 1-14.

Daniels, L. M. (2019). Objective score versus subjective satisfaction: Impact on emotions following immediate score reporting. International Journal of Experimental Education. https://doi.org/10. 1080/00220973.2019.1628692

Daniels, L. M., & Gierl, M. (2017). Emotions and immediate exam score reporting. Learning and Instruction, 52, 27–35.

Dweck, C. S. (2008). *Mindset: The new psychology of success*. Random House Digital, Inc.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational psychologist*, *34*(3), 169-189.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, *100*(3), 613-628.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels?. *Psychological bulletin*, *136*(3), 422-449

Isbell, L. M., Cote, N. G. (2009) Connecting with struggling students to improve performance in large classes. *Teaching of Psychology 36*(3): 185–188. doi: 10.1080/ 00986280902959960. https://journals.sagepub.com/doi/abs/10.1080/00986280902959960

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, *119*(2), 254-284.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of educational research*, *47*(2), 211-232.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational behavior and human decision processes*, *50*(2), 212-247.

Mouratidis, A., Michou, A., Demircioğlu, A. N., & Sayil, M. (2018). Different goals, different pathways to success: Performance-approach goals as direct and mastery-approach goals as indirect predictors of grades in mathematics. *Learning and Individual Differences*, *61*, 127-135.

Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and higher Education*, *31*(1), 53-64.

Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction*, *29*, 115-124.

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.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological bulletin*, *138*(2), 353-387.

Richardson, J. T. (1994). Mature students in higher education: Academic performance and intellectual ability. *Higher Education*, *28*(3), 373-386.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565-600.

Senko, C., & Harackiewicz, J. M. (2005). Regulation of Achievement Goals: The Role of Competence Feedback. *Journal of Educational Psychology*, *97*(3), 320–336.

Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. *Handbook of test development*, 329-347.

Wickline, V. B., & Spektor, V. G. (2011). Practice (rather than graded) quizzes, with answers, may increase introductory psychology exam performance. *Teaching of Psychology*, *38*(2), 98-101.

Table 1

Principal Component Analysis of Items Measuring Usefulness of Score Report Forms and Descriptive Statistics (n = 112)

	Percentage-only report	Descriptive feedback report	Choice score report (95% descriptive)
This feedback…	Factor loadings	Factor loadings	Factor loadings
encourages me to put in more effort.	.61	.77	.88
will help me reach my goals.	.85	.83	.92
is helpful.	.80	.86	.86
will help me perform better on the next exam.	.86	.85	.88
is the type I want after an exam.	.82	.84	.89
Eigenvalue	3.17	3.44	3.93
Percent of Variance	63.38	68.82	78.62
Coefficient alpha	.85	.87	.91

Table 2

Descriptive Statistics and Reliability of Measures

	Reliability	Range	Mean	Standard Deviation
Mastery-approach	.77	9-15	13.62	1.40
Mastery-avoidance	.89	3-15	11.45	3.38
Performance-approach	.90	4-15	12.10	2.63
Performance-avoidance	.92	3-15	11.46	3.29
Exam 1	--	46-98%	78.25%	9.76%
Exam 2	--	55-97%	81.76%	9.63%
Exam 3	--	48-93%	75.37%	9.40

Table 3

Zero-Order Correlation Matrix for All Variables

	1	2	3	4	5	6	7	8	9	10	11
1. Age											
2. Gender	-.19										
3. MAp	.08	.14									
4. MAv	.06	.14	.43*								
5. PAp	-.01	-.20	.32*	.36*							
6. PAv	-.04	.06	.24	.57*	.75*						
7. Exam 1 Score	.28*	-.01	.02	.02	-.01	-.13					
8. Usefulness 1	-.11	-.02	-.12	-.08	-.08	-.03	.06				
9. Exam 2 Score	.36*	-.04	.08	.04	.04	-.05	.68*	-.08			
10. Usefulness 2	.04	-.02	.03	.11	-.04	.01	.15	.31*	.23		
11. Exam 3 Score	.23	.07	.06	.04	.10	-.02	.61*	-.21	.68*	.08	
12. Usefulness 3	-.10	.10	-.04	.01	-.13	-.08	.11	.28*	.15	.47*	.07

* $p < .01$, Notes: Gender 1 = men 2 = women; MAp = Mastery-approach; MAv = Mastery-avoidance; PAp = Performance-approach; PAv = Performance-avoidance; Usefulness 1 = usefulness of the percentage-only report; Usefulness 2 = usefulness of the descriptive report; Usefulness 3 = usefulness of the chosen report (95% descriptive report)

Table 4

Regression Analyses Predicting Exam Score by Achievement Goals, Immediate Score Report Usefulness, and Previous Performance

	Exam 1	Exam 2			Exam 3		
Predictor variables	Step 1 β	Step 1 β	Step 2 β	Step 3 β	Step 1 β	Step 2 β	Step 3 β
Age	.29*	.36*	.19*	.17*	.26*	.01	-.01
Gender [a]	.13	.05	-.02	-.02	.22*	.18*	.17*
Mastery-approach	-.08	.01	.05	.05	-.09	-.05	-.05
Mastery-avoidance	.14	-.01	-.04	-.05	.06	.05	.06
Performance-approach	.31*	.25	.02	.002	.39*	.26*	.24
Performance-avoidance	-.42*	-.24	.03	.04	-.33	-.20	-.19
PE[b] performance	--		.61***	.62***		.66***	.68***
Usefulness of immediate score report from previous exam	--			-.10			-.06
Adjusted R^2	.08	.10	.44***	.44	.06	.45***	.45

* $p < .05$, *** $p < .001$ [a] 1 = men 2 = women, [b] PE = previous exam

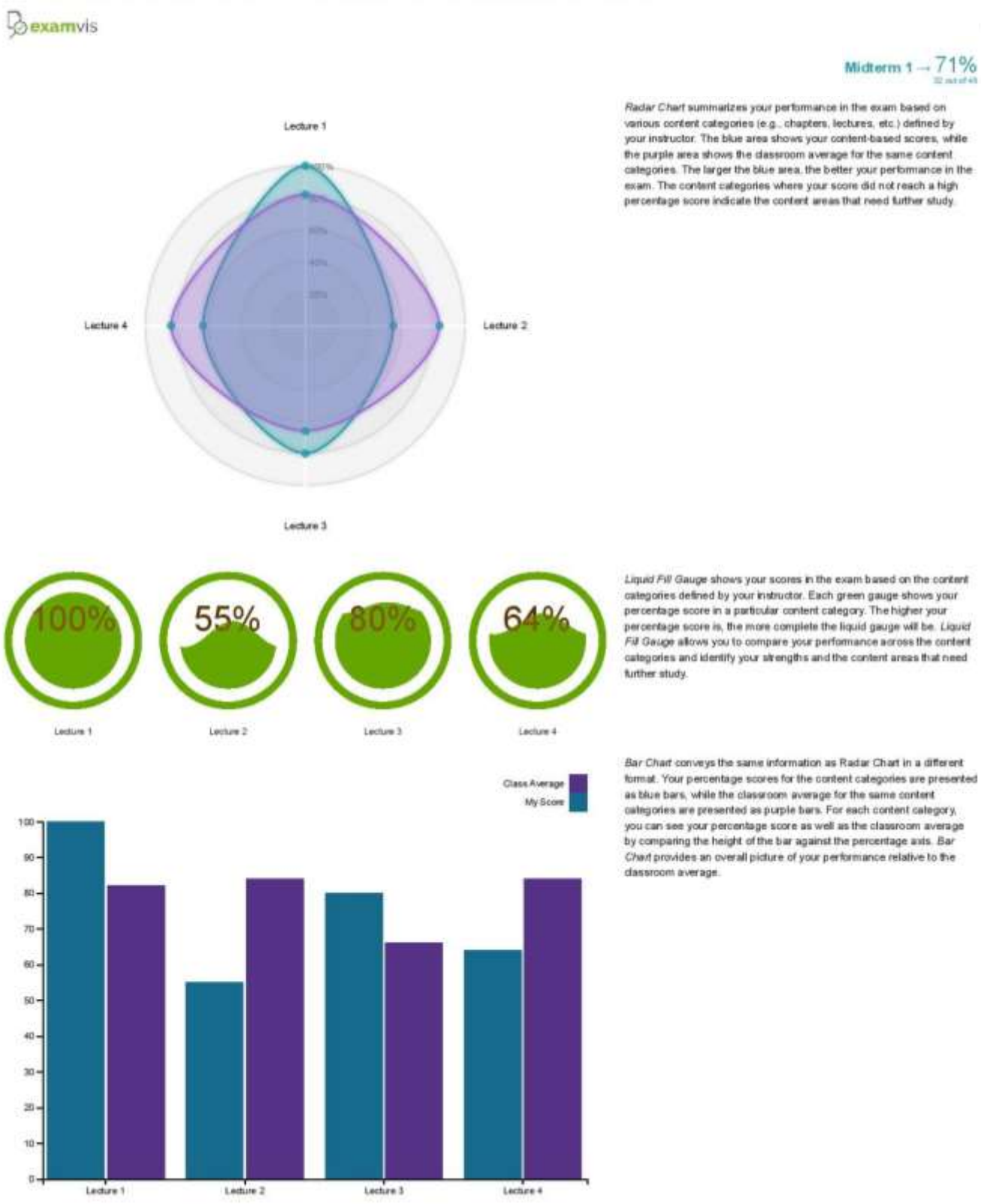

Figure 1. Visualizations in the sample descriptive feedback report.

Biological Development	You have a strong understanding of the biology behind adolescence. Consider building on this knowledge in your ePortfolio!
Cognitive Development	Your knowledge about cognitive development during adolescence is strong. This will be a great advantage when you move into the classroom!
School environment	You seem to have missed some key ideas about the influence of the school environment on student learning. The topics listed below highlight areas you may want to review especially before submitting your ePortfolio: School climate, class size, transitions, performance vs. mastery orientation
Culture and Technology	You struggled a bit with items on culture and technology. The topics listed below highlight areas you may want to review before you start your practicum: Ethnocentrism, media theories, Web 2.0 technologies

Figure 2. A sample of positive and constructive written feedback for each domain the descriptive feedback report