Abstract

Achievement goals are a powerful construct for understanding students’ classroom experiences and performance, yet most work examining achievement goals relies on self-report measures gathered through questionnaires. The current work aims to assess achievement goals using a task choice embedded within a typical classroom activity. Results show the behavioral measure of achievement goals predicts performance on the task, while self-reported achievement goals do not. Self-reported achievement goals predict quarterly grades, while the behavioral measure of achievement goals does not. This work supports the viability of a behavioral measure and suggests the achievement goals that students adopt at a task level may be different from their general class achievement goals. Using complementary achievement goal measures may improve understanding of how achievement goals relate to student behaviors and academic achievement.

Keywords: achievement goals; motivation; measurement; performance; learning

Introduction

In recent years, educational researchers have used achievement goals to predict behaviors and learning outcomes in academic environments (for reviews, see Hulleman, Schrager, Boddman, & Harackiewicz, 2010, and Linnenbrink-Garcia, Tyson, & Patall, 2008). Achievement goals are a construct defined by whether learners assess their competence using normative standards, i.e., judging achievement relative to others, or using intrapersonal standards, i.e., judging achievement relative to one’s self (Elliot & McGregor, 2001). These two types of goals have been labeled performance (normative) and mastery (intrapersonal) goals. Achievement goals are also defined by whether the learner strives to attain positive outcomes (approach goals) or evade negative outcomes (avoidance goals). Crossing these two dimensions produces four distinct goals: mastery approach, mastery avoidance, performance approach, and performance avoidance (Elliot & McGregor, 2001).

Performance-approach goals are associated with a mixture of positive behaviors such as effort and persistence, and negative feelings and behaviors such as shallow processing and test anxiety (Elliot, McGregor, & Gable, 1999; Elliot & McGregor, 2001). They positively predict classroom achievement measured by tests and grades in a subset of studies, while other studies have shown no relationship (Hulleman et al., 2010; Linnenbrink-Garcia, Tyson, & Patall, 2008). In contrast, performance-avoidance goals are consistently associated with negative feelings such as test anxiety; negative behaviors such as disorganization and shallow processing; and negative achievement outcomes (Elliot & McGregor, 2001; Elliot, McGregor, & Gable, 1999).

Mastery-approach goals have consistently been associated with greater interest, more enjoyment of the learning process, and productive learning behaviors such as deeper processing and persistence (Elliot, McGregor, & Gable, 1999). As with performance-approach goals, there is mixed evidence for their relationship to classroom achievement (i.e., tests and grades). Recent reviews of the literature have found positive associations with achievement for only a subset of the studies reviewed, while other studies have shown no relationship (Hulleman et al., 2010; Linnenbrink-Garcia, Tyson, & Patall, 2008). Like performance-approach goals, mastery-avoidance goals have been associated with positive behaviors including engagement and help-seeking as well as negative feelings and behaviors such as test anxiety, disorganization, and surface processing (Elliot & McGregor, 2001).

Most of the achievement goal data collected in classroom settings has relied on self-report questionnaires (Linnenbrink-Garcia, Tyson, & Patall, 2008). Although many achievement goal scales are reliable (Hulleman et al., 2010), self-report measures are subject to many
limitations including issues of context effects, goal specificity (e.g., domain- vs. task-level), timing of the goal assessment, external validity, and difficulties in accurately assessing one’s own cognitions. Although questionnaires have provided a great deal of insight into the predictive value of achievement goals for performance and learning outcomes (e.g., tests, grades, SATs), the creation of a behavioral measure (e.g., task choice, strategy selection) could improve understanding of how achievement goals relate to student behaviors, classroom goal adoption, and academic achievement (Fulmer & Frijters, 2009). A behavioral measure could also help to clarify the inconsistent performance outcomes associated with mastery-approach and performance-approach goals (e.g., Linnenbrink-Garcia, Tyson, & Patall, 2008).

The current work assesses a behavioral measure that infers achievement goals through a task choice embedded within a typical classroom activity. After reviewing the literature supporting the need for such a measure, we present the relationships between our behavioral data and students’ performance on an embedded task, their broader content knowledge, and their self-reported achievement goals. Finally, we explore how this measure could deepen our understanding of the performance and learning outcomes associated with different achievement goals.

**Theoretical framework**

There have been two distinct approaches to achievement goal research (Barron & Harackiewicz, 2001). The first takes a dispositional approach by assessing goals using self-report measures, typically at the domain level in classroom settings, and assumes that achievement goals are relatively stable over time (e.g., Elliot & McGregor, 2001). The second is a more dynamic perspective that views goals as subject to quick changes and manipulation through experiments and classroom interventions (e.g., Elliot & Dweck, 1988). In the present study, we investigate the potential value of both perspectives in predicting outcomes, as well as the possibility that they measure different things. Although we do not question the well-established internal reliability of self-reported achievement goals or their myriad relationships to the aforementioned achievement outcomes, there may be a gap between questionnaire responses and the goal-directed behaviors students demonstrate when engaging in classroom activities.

**Self-reported goal measures**

Despite the significant contributions of data from self-report questionnaires to our understanding of student motivation, there are several disadvantages of such an approach that suggest a need for complementary measures (Filmer & Frijters, 2009). First, there may be a misalignment between the context targeted by a questionnaire and achievement goals activated during the tasks in which students engage during a course. If individual, task, domain, and environmental contexts all play roles in determining which achievement goals are activated at a given moment (Pintrich, 2003), then achievement goal measures that are distinctly separate from typical instructional tasks (e.g., completing a questionnaire before or after class) might reflect different goals compared to measures administered within the context of typical classroom activities. There may also be a misalignment in goal specificity, such that the goals students have for individual tasks might differ from the goals they have for the course or domain. Students may not understand the meaning of items, be sensitive to differences among items, or carefully consider the options when responding, any of which could decrease the external validity of the measure. Furthermore, students’ goals change throughout the course of the semester (Fryer & Elliot, 2007), and it is unclear what time scales are appropriate when relating achievement goal measures to measures of other variables such as feelings, behaviors and outcomes. Finally, self-report measures assume that goals are consciously accessible, which contradicts evidence that many people have poor metacognitive awareness (Metcalfe, Eich, & Castel, 2010).

**Behavioral goal measures**

Several lines of research have recently used behavioral data captured in computer learning environments to infer achievement goals. In a laboratory setting, Zhou and Winne (2012) gave participants a multimedia-enhanced article to study, instructing them that they could use tags to take notes on the article or follow hyperlinks to additional information to enhance their learning. The information built into the tag system and the hyperlinks was aligned to the four achievement goal constructs (e.g., a mastery-approach tag stated, “I want to learn more about this,” while a mastery-approach hyperlink was labeled, “Find more information about this”). The authors tracked how frequently participants used tags and hyperlinks aligned with each of the four constructs to create behavioral measures of participants’ goals. Achievement goals inferred through behavioral traces were not associated with self-reported achievement goals. The authors then used individual achievement goals to predict performance on a test about the content of the article. In each regression model, a behavioral goal was entered first and its corresponding self-reported goal was entered second (e.g., behavioral mastery-approach goal at step one and self-reported mastery-approach goal at step two). Behavioral measures of mastery approach, performance approach, and performance avoidance were all positive predictors of performance, and no self-reported goal explained any additional variance in its respective model.

This work provides an important first step toward assessing behavioral indicators of achievement goals, and it suggests that behaviorally inferred achievement goals might be better predictors of task performance than self-reported goals. However, as each goal was analyzed in
isolation from the other goals and all four behavioral goals were significantly, positively correlated to task performance, it is unclear that participants’ use of different tags or hyperlinks represented distinct goals. Given the fairly consistent past findings relating to performance-avoidance goals, it is particularly surprising that performance-avoidance goals were associated with positive performance outcomes. Since the behavioral traces from which these goals were inferred reflected participants’ highlighting of text and seeking additional information through hyperlinks, it may be that such behavioral traces reflected general study strategies or engagement with the text. Additionally, participants may have had little context to guide their responses to the self-report questionnaire in a laboratory setting.

Attempting to replicate Zhou and Winne’s (2012) findings using data from an intelligent tutoring system, Otieno, Schwonke, Salden, and Renkl (2013) assessed traces of students’ behaviors in a classroom setting. The authors used students’ access of hints, which they characterized as reflecting a focus on solving problems, as an indicator of performance-approach goals and students’ access of glossary terms, which they characterized as reflecting their focus on understanding principles, as an indicator of mastery-approach goals. Like Zhou and Winne (2012), they found no significant correlations between self-reported goals and their behavioral measures of goals. However, they found that their two behavioral goal measures were strongly, negatively correlated. Hint use (indicating performance-approach goals) was negatively associated with performance on a series of principle-based post-tests while glossary use (indicating mastery-approach goals) was positively associated with immediate post-test performance. Self-reported goals were generally less predictive of performance, with the exception of self-reported mastery-approach goals predicting delayed post-test performance. This work addresses some of the limitations in Zhou and Winne’s (2012) study by assessing goals in a classroom context and clearly differentiating between behavioral measures, but the connections between learners’ use of resources and the constructs targeted by self-reported achievement goal questionnaires are less clear. Further, the relationship between goals contradicts most findings using self-reported measures; while Otieno et al. (2013) found mastery-approach and performance-approach goals to correlate negatively, most past work has found no correlation or a moderate, positive correlation between the two (Barron & Harackiewicz, 2011).

The Present Study

Given Pintrich’s (2003) argument for the role of context in activating achievement goals, a behavioral measure administered in the context of an academic task should more accurately reflect the achievement goals students experience when engaged in class activities. In the present study, we assessed achievement goals through a goal-framed task choice embedded in a classroom activity. Task choice has been used more broadly to assess other motivational constructs (e.g., Atkinson, 1964), and it has been predicted using a two-dimension achievement motivation framework (Nicholls, 1984). Students’ task choice selections should reflect their achievement goals as they relate to achievement on the task itself, permitting an examination of the relationship between self-reported achievement goals, task choice, task outcomes, and class grades. We tested the following hypotheses:

(H1) Self-reported goals will weakly predict task-based goals, as these two constructs are related but are being assessed at different levels (task vs. domain) and in different contexts.

(H2) Task-based goals will predict task performance better than self-reported goals, as a result of task choice occurring in a more similar context to the task completion and assessment.

(H3) Self-reported goals will predict grades, which are a reflection of many different activities and choices amassed over time.

Methods

Participants

One hundred and one students from four science classes at an urban, public middle school participated in the study. Students were enrolled in two seventh grade (51 students) and two eighth grade classes (50 students), with a different teacher for each grade. Participation occurred as part of regular classroom activities, with students receiving participation credit for completing the materials. Twenty-five students (2 seventh graders, 23 eighth graders) were absent on the day the task-based measure was administered or did not complete the task materials and were therefore excluded from analyses examining task goals or task performance (remaining \( n = 76 \)). Sixteen students (8 seventh graders, 8 eighth graders) were absent on the day the self-report measure was administered or did not complete it and therefore were excluded from analyses examining self-reported achievement goals (remaining \( n = 85 \)). For analyses relating self-reported achievement goals to task achievement goals, a total of 38 students (12 seventh graders, 26 eighth graders) were excluded because they were missing one or both measures (remaining \( n = 63 \)).
Performance Avoidance  Packet C  The third stack contains an activity that will focus on helping you avoid performing worse than other students typically perform in this unit. You should choose this if you want to try to avoid performing worse than other students.

Materials
Our primary behavioral measure of achievement goals took the form of students' choices among three sets of activities labeled Packet A, Packet B, and Packet C (Table 1). The descriptions for each packet appeared on a teacher-read script, a slide projected at the front of class, and on the cover page of the packets themselves. Achievement goal labels were not present in the descriptions. The choices posed by our behavioral measure were aligned to Elliot and Murayama's (2008) Achievement Goals Questionnaire-Revised (AGQ-R), with options corresponding to constructs of mastery approach, performance approach, and performance avoidance. We excluded mastery avoidance in response to younger students' difficulty in conceptualizing its meaning. Aside from the cover pages stating the descriptions, all packets contained identical materials. Each teacher selected a popular science article about a curriculum-appropriate topic to be used in the task. The packets also included five comprehension questions to be completed after students read the article.

Additional measures included students' performance on the packet comprehension questions; their quarterly content grades, which were based on quiz, test, homework, and project scores; and their responses on 7-point Likert scales to the 12-item (three per construct) AGQ-R framed around science class.

Procedure
Three days prior to the administration of the task-based measure, students completed the AGQ-R during science class. On the day of the task-based measure, students were told they would complete a reading comprehension task in class. They were asked to choose among three versions of the task based on their goals. To decrease pressure to be seen choosing a particular packet by their peers, and to prevent students from examining the packets before making their choices, they were told to write down their choices. The teachers then handed out packets based on the choice each student indicated.

Results
Analyses focused on the relationship between students' packet choices and responses on the AGQ-R (H1), the comparative strength of the relationship between performance on the packets and the two goal measures (H2), and the comparative strength of the relationship between end-of-quarter grades and the two goal measures (H3). We report effect sizes (R² or partial eta squared, n²p) for all significant main effects, and we interpret effects as small when $R²$ or $n²p < .06$, medium when $.06 < R²$ or $n²p < .14$, and large when $R²$ or $n²p > .14$.

Self-reported goals and task choice
To assess the relationship between packet choice and students' self-reported endorsements of mastery approach, performance approach, and performance avoidance on the AGQ-R, packet choice was recoded into dichotomous variables for each choice (e.g., coded as either “chose mastery packet” or “did not choose mastery packet”), and logistic regression was used to assess whether self-reported goals predicted the likelihood of choosing a particular packet. Self-reported mastery-approach-performance-approach, and performance-avoidance goals were all included in the models; mastery avoidance was excluded from all regressions throughout the analyses because there was no corresponding option among the packet choices.

The logistic regression model predicting likelihood of choosing a particular packet was not significantly better than a constant-only model for mastery-approach packets, $X²(3, N = 63) = 2.42, p = .49$, performance-approach packets, $X²(3, N = 63) = 3.89, p = .27$, or performance-avoidance packets, $X²(3, N = 63) = 4.74, p = .19$. Although we must use caution interpreting a model that is not significant, we examined the models for significant variables and found only a marginally significant variable, self-reported performance-approach goal endorsement, Wald's $\chi²(1, N = 63) = 3.14, p = .08, \text{Exp (B) } = 2.92$, within the model predicting performance-approach packet selection. These results suggest performance approach as measured by the AGQ-R may weakly predict performance-approach task choice whereas the AGQ-R mastery-approach and performance-avoidance measures were not related to their behavioral counterparts. These results are generally consistent with previous findings relating behaviorally inferred achievement goals to self-reported achievement goals (Zhou & Winne, 2012; Otieno et al., 2013).

### Table 1: Framing of packet choices for task-based behavioral measure of achievement goals

<table>
<thead>
<tr>
<th>Achievement goal</th>
<th>Label</th>
<th>Description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mastery Approach</td>
<td>Packet A</td>
<td>The first stack contains an activity that will focus on helping you improve your understanding of the concepts in this unit. You should choose this if you want to try to completely understand the concepts in this unit.</td>
<td></td>
</tr>
<tr>
<td>Performance Approach</td>
<td>Packet B</td>
<td>The second stack contains an activity that will focus on helping you perform better than other students typically perform in this unit. You should choose this if you want to try to perform better than other students.</td>
<td></td>
</tr>
<tr>
<td>Performance Avoidance</td>
<td>Packet C</td>
<td>The third stack contains an activity that will focus on helping you avoid performing worse than other students typically perform in this unit. You should choose this if you want to try to avoid performing worse than other students.</td>
<td></td>
</tr>
</tbody>
</table>

### Materials

Our primary behavioral measure of achievement goals took the form of students' choices among three sets of activities labeled Packet A, Packet B, and Packet C (Table 1). The descriptions for each packet appeared on a teacher-read script, a slide projected at the front of class, and on the cover page of the packets themselves. Achievement goal labels were not present in the descriptions. The choices posed by our behavioral measure were aligned to Elliot and Murayama's (2008) Achievement Goals Questionnaire-Revised (AGQ-R), with options corresponding to constructs of mastery approach, performance approach, and performance avoidance. We excluded mastery avoidance in response to younger students' difficulty in conceptualizing its meaning. Aside from the cover pages stating the descriptions, all packets contained identical materials. Each teacher selected a popular science article about a curriculum-appropriate topic to be used in the task. The packets also included five comprehension questions to be completed after students read the article.

### Procedure

Three days prior to the administration of the task-based measure, students completed the AGQ-R during science class. On the day of the task-based measure, students were told they would complete a reading comprehension task in class. They were asked to choose among three versions of the task based on their goals. To decrease pressure to be seen choosing a particular packet by their peers, and to prevent students from examining the packets before making their choices, they were told to write down their choices. The teachers then handed out packets based on the choice each student indicated.

### Results

Analyses focused on the relationship between students' packet choices and responses on the AGQ-R (H1), the comparative strength of the relationship between performance on the packets and the two goal measures (H2), and the comparative strength of the relationship between end-of-quarter grades and the two goal measures (H3). We report effect sizes ($R²$ or partial eta squared, $n²p$) for all significant main effects, and we interpret effects as small when $R²$ or $n²p < .06$, medium when $.06 < R²$ or $n²p < .14$, and large when $R²$ or $n²p > .14$.

### Self-reported goals and task choice

To assess the relationship between packet choice and students' self-reported endorsements of mastery approach, performance approach, and performance avoidance on the AGQ-R, packet choice was recoded into dichotomous variables for each choice (e.g., coded as either “chose mastery packet” or “did not choose mastery packet”), and logistic regression was used to assess whether self-reported goals predicted the likelihood of choosing a particular packet. Self-reported mastery-approach-performance-approach, and performance-avoidance goals were all included in the models; mastery avoidance was excluded from all regressions throughout the analyses because there was no corresponding option among the packet choices.

The logistic regression model predicting likelihood of choosing a particular packet was not significantly better than a constant-only model for mastery-approach packets, $X²(3, N = 63) = 2.42, p = .49$, performance-approach packets, $X²(3, N = 63) = 3.89, p = .27$, or performance-avoidance packets, $X²(3, N = 63) = 4.74, p = .19$. Although we must use caution interpreting a model that is not significant, we examined the models for significant variables and found only a marginally significant variable, self-reported performance-approach goal endorsement, Wald's $\chi²(1, N = 63) = 3.14, p = .08, \text{Exp (B) } = 2.92$, within the model predicting performance-approach packet selection. These results suggest performance approach as measured by the AGQ-R may weakly predict performance-approach task choice whereas the AGQ-R mastery-approach and performance-avoidance measures were not related to their behavioral counterparts. These results are generally consistent with previous findings relating behaviorally inferred achievement goals to self-reported achievement goals (Zhou & Winne, 2012; Otieno et al., 2013).
Predicting task performance

Overall, students demonstrated a preference for the mastery-approach packet ($N = 40$) over the performance-approach packet ($N = 17$) and the performance-avoidance packet ($N = 19$). These frequencies were significantly different, $X^2(2, N = 76) = 12.82, p < .01$. This distribution of responses might seem to suggest students were inclined to choose the packet they considered most desirable to the teacher, or the first packet in the list; however, this pattern is consistent with past work showing the majority of students express a dominant mastery goal (Van Yperen, 2006). Furthermore, if packet choice were based on desirability or list order, we would not expect it to predict task performance. Performance was measured by the number of correctly answered questions from the reading and ranged from zero to five. A one-way analysis of variance (ANOVA) revealed packet choice had a large effect on performance on the packet comprehension questions, $F(2,72) = 8.61, p = .00, n_p^2 = .19$, with students who chose the mastery-approach packet ($M = 3.05, SD = 1.61$) and the performance-approach packet ($M = 3.71, SD = 1.00$) scoring better than those who chose the performance-avoidance packet ($M = 1.89, SD = .94$).

By comparison, a multivariate regression analysis predicting task performance with students’ self-reported mastery-approach, performance-approach, and performance-avoidance goals indicated the three AGQ-R predictors explained none of the variance ($R^2 = .04, F(3,62) = .77, p = .52$). Within the model, no goals were significantly predictive of performance (Table 2).

While achievement goals indicated by packet choice were highly predictive of performance on the subsequent task, students’ self-reported achievement goals did not predict performance. This suggests the AGQ-R might not reflect specific task achievement goals as accurately as achievement goals inferred through a behavioral measure incorporated in the task.

### Table 2: Summary of multivariate regression analysis for AGQ-R ratings predicting task performance

<table>
<thead>
<tr>
<th>Achievement goal</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mastery Approach</td>
<td>-.168</td>
<td>.190</td>
<td>-.156</td>
</tr>
<tr>
<td>Performance Approach</td>
<td>.315</td>
<td>.309</td>
<td>.288</td>
</tr>
<tr>
<td>Performance Avoidance</td>
<td>-.028</td>
<td>.263</td>
<td>-.026</td>
</tr>
</tbody>
</table>

*p < .05

Predicting grades

A one-way ANOVA indicated packet choice was not associated with students’ content grades for the quarter during which they completed the task, $F(2,76) = 1.77, p = .18$. By comparison, a multivariate regression analysis predicting content grade with students’ self-reported achievement goals indicated the three AGQ-R predictors explained 20.1% of the variance ($R^2 = .23, F(3,81) = 8.03$, $p < .01$). Within the model, only performance approach was significantly predictive of content grade (Table 3).

These results are consistent with work demonstrating self-reported achievement goals for a domain can predict grades. The difference in predictive power between our behavioral measure and the AGQ-R might stem from framing the measure around a specific task (behavioral measure) versus a domain (AGQ-R). The results suggest achievement goals indicated by packet choice do not scale up to the domain level and predict grades.

### Discussion

Although more research must be conducted to assess the strength of these findings, this work demonstrates the difference between context-based, task-specific achievement goals and more stable, domain-general achievement goals, as well as the value of each for predicting different levels of achievement. Self-reported achievement goal measures have made significant contributions to the field of motivation research, but they are subject to limitations stemming from context, specificity, timing, external validity, and students’ metacognitive skills. Researchers should continue investigating alternative measures that can be integrated into achievement tasks. Very little research examines the extent to which achievement goals can be inferred from behavioral data, although important first steps are being made (Otieno et al., 2013; Zhou & Winne, 2012).

The data suggest our behavioral measure can predict achievement at the task level on which the achievement goals were measured. Relationships between task-framed, behavioral measures and domain-framed, self-reported measures must be further explored, but the absence of correlation is consistent with limited prior work on this issue. Nearly half of all students selected a mastery-approach packet. While this is consistent with past work on dominant achievement goals (Van Yperen, 2006), it is also possible that students were biased toward selecting the mastery-approach packet because it was the first option listed. Future work should randomize the order of packet choices across classes to see if this bias persists.

Students were permitted to select only one packet, creating a dichotomous assessment of achievement goals that cannot capture the multiple goals many students possess. However, this framework forces students to act on the dominant achievement goal activated in the task context. While students might endorse multiple self-reported achievement goals with equal strength, this forced-choice design may better capture the achievement
goal most relevant to a student at a particular moment. Repeated measures could provide a richer picture of students’ goal orientations as they relate to different contexts (see Bernacki, Nokes-Malach, & Aleven, 2013).

There was no significant relationship between goal choice and self-reported goals. Given that more than a third of students were excluded from the analysis due to missing data, and that task choice was not equally distributed across the three options, a larger sample size might be necessary to detect significant relationships between self-reported goals and packet choice. The failure for self-reported measures to predict packet choice might reflect students’ inability to consciously access their goals, or it might be a result of the domain framing of the AGQ-R (i.e., students self-reported their goals for science class, and not for a specific task in science class). If the latter were the case, we would expect a model predicting task choice with responses to a task-framed AGQ-R to yield greater reliability. Future work should examine the different scales of achievement goals as task-based versus domain-based and explore how data from this task-based, behavioral achievement goal measure might change from task to task and across the academic year. The differences between Likert scale responses (AGQ-R) and a forced choice (task choice) might also explain the lack of relationship between measures.

Better understanding the relationship between self-reported achievement goals and behaviorally inferred achievement goals is of particular importance. If behavioral measures assumed to indicate students’ pursuits of achievement goals consistently fail to correlate with self-reported assessments of achievement goals, we must question whether such behavioral measures reflect the same constructs as the self-report assessments.

Acknowledgments
This research was supported by Grant SBE00354420 from the National Science Foundation to the Pittsburgh Science of Learning Center (http://www.learnlab.org). We thank teachers Paul Ronevich and Brian Rose for their insightful feedback and assistance in developing materials and conducting this research.

References
Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. Learning and Instruction, 22, 413-419.