Summer 2017

ASPECT: A Survey to Assess Student Perspective of Engagement in an Active-Learning Classroom

Benjamin L. Wiggins  
*University of Washington*

Sarah L. Eddy  
*Florida International University*, sarah.eddy@fiu.edu

Leah Wener-Fligner  
*University of Washington*

Karen Freisem  
*University of Washington*

Daniel Z. Grunspan  
*Arizona State University*

See next page for additional authors

Follow this and additional works at: [https://digitalcommons.unl.edu/sociologyfacpub](https://digitalcommons.unl.edu/sociologyfacpub)

Part of the [Family, Life Course, and Society Commons](https://digitalcommons.unl.edu/familylifecourseandsoccom), and the [Social Psychology and Interaction Commons](https://digitalcommons.unl.edu/socialpsychologyandinteractioncom)

---

Wiggins, Benjamin L.; Eddy, Sarah L.; Wener-Fligner, Leah; Freisem, Karen; Grunspan, Daniel Z.; Theobald, Elli J.; Timbrook, Jerry; and Crowe, Alison J., "ASPECT: A Survey to Assess Student Perspective of Engagement in an Active-Learning Classroom" (2017). *Sociology Department, Faculty Publications*. 497.  
[https://digitalcommons.unl.edu/sociologyfacpub/497](https://digitalcommons.unl.edu/sociologyfacpub/497)

---

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Sociology Department, Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
ASPECT: A Survey to Assess Student Perspective of Engagement in an Active-Learning Classroom

Benjamin L. Wiggins,† Sarah L. Eddy,† Leah Wener-Fligner,† Karen Freisem,*, Daniel Z. Grunspan,*, Elli J. Theobald,† Jerry Timbrook,** and Alison J. Crowe**

†Department of Biology, University of Washington, Seattle, WA 98195-1800; †Biology Department and STEM Transformation Institute, Florida International University, Miami, FL 33199; *College of Education, University of Washington, Seattle, WA 98195-3600; **Center for Teaching and Learning, University of Washington, Seattle, WA 98195-1265; #Center for Evolution and Medicine, Arizona State University, Tempe, AZ 85287-4501; **Department of Sociology, University of Nebraska, Lincoln, NE 68588

ABSTRACT
The primary measure used to determine relative effectiveness of in-class activities has been student performance on pre/posttests. However, in today’s active-learning classrooms, learning is a social activity, requiring students to interact and learn from their peers. To develop effective active-learning exercises that engage students, it is important to gain a more holistic view of the student experience in an active-learning classroom. We have taken a mixed-methods approach to iteratively develop and validate a 16-item survey to measure multiple facets of the student experience during active-learning exercises. The instrument, which we call Assessing Student Perspective of Engagement in Class Tool (ASPECT), was administered to a large introductory biology class, and student responses were subjected to exploratory factor analysis. The 16 items loaded onto three factors that cumulatively explained 52% of the variation in student response: 1) value of activity, 2) personal effort, and 3) instructor contribution. ASPECT provides a rapid, easily administered means to measure student perception of engagement in an active-learning classroom. Gaining a better understanding of students’ level of engagement will help inform instructor best practices and provide an additional measure for comprehensively assessing the impact of different active-learning strategies.

INTRODUCTION
National reports aimed at improving undergraduate science education have called for a shift away from the traditional “sage on a stage” mode of lecturing toward the use of student-centered, evidence-based instructional approaches (National Research Council, 2003; American Association for the Advancement of Science, 2011; President’s Council of Advisors on Science and Technology, 2012). This is due in part to the fact that increasing the amount of active learning in the classroom has been shown to benefit student learning (Freeman et al., 2014). In an active-learning environment, students spend more time coconstructing knowledge with their peers (Chi and Wylie, 2014), which requires the ability to form effective working interactions with friends or peer strangers (Lorenzo et al., 2006). Many factors have been found to influence whether or not students actively engage in small-group work, including English language proficiency, perceived value of the activity, and group composition (Chatman et al., 2008; Dasgupta and Stout, 2014; Grunspan et al., 2014). However, there is little literature on how more social rather than individual learning is impacting students’ experience in the classroom (Kurth et al., 2002; Hand, 2006). For example, being the only member of a particular social category (e.g., gender or ethnicity) has the potential...
Engagement is a multifaceted concept and includes dimensions ranging from behavioral (being on task) to cognitive (exerting effort) to affective (being invested in a task) (Christenson et al., 2012; Reeve and Lee, 2014). As our interest is in how students engage with active-learning exercises, we will use the term “engagement” here to mean “learning task engagement” as defined by Chapman to encompass “students’ cognitive investment, active participation, and emotional engagement with specific learning tasks” (Chapman, 2003, p. 1). Many studies have shown a positive correlation between student engagement and achievement (Dweck, 1986; Wigfield and Eccles, 2000; Hidi and Renninger, 2006; Hulleman et al., 2008; Chi and Wylie, 2014; Reeve and Lee, 2014), leading to the development of a number of different theoretical frameworks to explain this relationship (Dweck, 1986; Wigfield and Eccles, 2000; Chi and Wylie, 2014). Although the underlying motivations driving engagement may vary, it is clear that measuring the extent to which students do or do not engage is important for comprehensive assessment of the effectiveness of active-learning strategies.

There are already several classroom observation tools that can be used to measure overall student participation in the classroom (Sawada et al., 2002; Hora and Ferrare, 2010; Smith et al., 2013; Eddy et al., 2015). An additional observation tool was recently developed to specifically assess student behavioral engagement in large college classrooms (Lane and Harris, 2015). However, these observation tools are limited (by design) to measuring overt behaviors and thus do not capture the internal level of investment or value students are placing on an activity. This can be problematic, according to Pritchard (2008), who documented poor correlation between outward manifestations of traditional “engaged” behaviors (such as sitting upright or looking at the instructor) and student self-reported engagement. This suggests that relying solely on classroom observation may not provide a complete picture of a student’s level of involvement. These findings are not surprising, as attentiveness can have many manifestations that fall outside “engaged” behavioral norms, so a student may be deeply engaged in a thought-provoking activity but not exhibit overt signs of engagement (Chi and Wylie, 2014). Ultimately, it is difficult to measure behavioral engagement and even more difficult to measure cognitive and affective engagement through external observation unless student work and attitudes are analyzed (Hart, 1994; Radford et al., 1995).

Alternatively, it is possible to assess student cognitive and affective engagement by asking students to reflect on their own levels of engagement. Several published questionnaires rely on self-report data to provide a more complete view of engagement (Chapman, 2003; Handelsman et al., 2005; Pazos et al., 2010). Although self-report data have the limitation that students may not accurately assess their own levels of engagement (Assor and Connell, 1992), it has the advantage of being able to provide some insight into why students find an activity more engaging, not just whether or not they are visibly engaged. However, of the surveys intended for college students, many are focused on a single aspect of a student’s experience such as personal motivation or sense of belonging (Pintrich et al., 1993; Hagerty and Patusky, 1995). Others are geared toward assessing student engagement in a traditional, lecture-based classroom (Handelsman et al., 2005) or are specific for a single type of active-learning strategy such as problem-based learning (Pazos et al., 2010).

Our goal here was to develop a more broadly applicable survey that would enable comparison of the relative effectiveness of different in-class activities at engaging students across the cognitive and affective dimensions of engagement. As we explain later, we have taken a mixed-methods approach to develop a survey that is grounded in the experience of undergraduate biology students and can be used to assess multiple aspects of student self-reported engagement. The survey was designed to be able to capture student engagement for a wide variety of active-learning strategies commonly used in college classrooms. We have chosen to focus on measuring students’ self-perception of engagement rather than measuring student behavior in order to capture the cognitive and affective dimensions of engagement. The survey is based on themes that arose during student interviews and focus groups and has been validated in a large introductory biology classroom. The resulting 16-item survey, which we call the Assessing Student Perspective of Engagement in Class Tool (ASPECT), can be used to rapidly obtain quantitative data on student self-reported engagement in an active-learning classroom.

**METHODS**

**Participants**

The students who participated in this study were enrolled in one of three quarters of an introductory biology course at a large research university in the Pacific Northwest. The course is the second course of a three-course series, with class size ranging from 370 to 760, depending on the quarter being taught. Different quarters of this course were taught by different instructors, but always included high levels of active learning, including clicker questions, group worksheets, case studies, peer instruction, and whole-class discussions. As described by registrar statistics, students enrolled in this course, over all three quarters, were primarily sophomores (49%) and juniors (40%) and had declared a wide range of majors, typically in the natural sciences. Female students made up on average 60% of the classroom population. In addition, of the students enrolled in the course, 44.2% were Asian Americans, 39.5% were white Americans, 6.3% were international, 5.5% were Latin@s, 1.8% were Black Americans, 1.8% were Hawaiian and Pacific Islander, and 0.8% were Native Americans. Community college transfer students made up 6% of the class, and 46% of the students were first-generation college students.

**Overview of Survey Development and Validation**

Here, we provide an overview of our survey development process (Figure 1), which follows the process described by Corwin et al. (2015) and is consistent with Benson’s validation framework (Benson, 1998). We have organized our description of the
development and validation of ASPECT into three phases that parallel the three stages established by Benson (1998): phase I, in which we 1) develop the constructs to be assessed, 2) design the survey items, and 3) obtain face validity for the survey items; phase II, in which we assess the dimensionality and reliability of the survey; and phase III, in which we gather evidence for the external validity of the survey. Each phase is described in more detail in Results.

Phase I: Development and Validation of Constructs Measured in ASPECT
In phase I, we conducted student interviews and focus groups to identify the constructs, or themes, of engagement on which to base the survey. Participants were recruited through blind carbon-copied email forwarded by the instructor to randomly chosen students in the class. The response rate averaged across all groups was 5–7%. In total, 25 participants were recruited into a series of interviews (n = 2) and focus groups (n = 7, ranging from two to five students per focus group) over the course of Fall 2012 and Winter 2013. Participants encompassed qualitatively similar characteristics to the class as a whole in terms of ethnicity, race, gender, and final course grades. In the interviews, we asked students general questions to elicit their thinking about the activity that had taken place in class that day. After transcription and coding, the number of lines of text was used as an (imperfect) approximation of frequency of each code within the transcript. The interview process and student themes arising from the interviews are described in Results.

From the themes arising out of the focus groups, we wrote Likert-scale items aimed at determining the overall engagement students experienced in class. Items were edited extensively based on student think-alouds and best practices of survey design (Dillman et al., 2014). The process of question development and revision is illustrated for one question (Figure 2) and included 1) standardizing the number of response alternatives to a six-point Likert scale across all items; 2) separating questions identified as containing two different ideas (i.e., “double-barreled”) into two distinct items to ensure that respondents were only asked about one idea per survey item; and 3) revising questions identified as having ambiguous wording to contain more explicit, straightforward language. In addition, several of the original items required students to compare their experiences during an intensive active-learning day with those of a “normal class day.” To remove possible confusion or alternative interpretations of a “normal class day,” we replaced these items with questions asking students to reflect directly on that day’s class (Figure 2). Finally, the survey was shortened to 20 items by removing redundant items (as determined through student think-alouds to be measuring the same general concept). Cognitive testing with a focus group of undergraduate biology students (n = 6) was performed to ensure the survey wording was clear and that students understood the intended meaning of each question. To obtain large-scale face validation of the survey items, we asked students to complete the entire survey online and then explain their thinking in open-ended responses for two to three randomly assigned survey items. This resulted in 30–40 short-answer responses per item. The process of coding these responses is described in Results.

Phase II: Validity and Reliability of ASPECT
In phase II, we assessed the dimensionality and reliability of the survey in three steps. First, we used pairwise correlation analysis to determine interitem correlation (Spearman’s rank correlation coefficient). Second, we used iterative exploratory factor
An example question matures through several steps:

Coding of original interviews identified ‘confidence’ as a key emergent theme in active learning classrooms. Initial survey item writing produced the following:

Version 3: Consider how confident that you are in your understanding of the material presented today. After today’s activity i am... a regular class day.
- Much more confident than
- More confident than
- Just as confident as
- Less confident than
- Much less confident than

Student think-alouds indicated four problems and/or multiple interpretations with this wording.

Version 4: I felt LESS confident in my understanding of the material covered after the activity today than after a regular class day

Better match to student language

Less reading load

Agree

Disagree

Clarified by positive phrasing

Less wordy (lower cognitive load)

Iterative editing produced two more versions, the second of which showed no problems for the participants interviewed.

Version 6: I felt MORE confident with this material.
- Agree
- Disagree

Clarified into a single construct

Standardized choices across survey items

Coding of 35 open-ended explanations revealed that no students varied from the intended interpretation, but students did include different reasons for their different confidence levels, primarily based either on procedural issues with the activity or their own level of preparedness.

FIGURE 2. Example of development process for one survey item. This question was iteratively improved through the qualitative steps discussed in Methods. Examples of specific changes in the development of this question are noted.

analysis (EFA) to determine the dimensionality of the survey and assess the internal consistency of the scales. For this, we used an oblique (promax) rotation, as we hypothesized that different aspects of engagement would be correlated with one another. We performed this EFA analysis on survey responses from online administration of ASPECT in a single quarter of introductory biology (n = 425). Students with missing responses (n = 17; 4.5% of total) were excluded from this analysis. All EFAs were conducted using the “psych” package in R (Revelle, 2014). Finally, to test the reliability and internal consistency of the scales identified by EFA, we used Cronbach’s alpha (Cronbach, 1951) as a measure of the internal consistency of the interrelatedness of the items. In both cases, EFA identified three factors: value of group activity (Value), personal effort (PE), and instructor contribution (IC).

Phase III: External Validity of ASPECT

In phase III, the final stage, we assessed whether ASPECT could discriminate between different activity types and different demographic populations as a measure of external validity of the survey. We compared student responses after completing either 1) a long-activity day in which students worked in groups to complete a worksheet (~30 minutes long) or 2) a short-activity day with a series of clicker-question activities centered around instructor-posed questions. Cronbach’s alpha (Cronbach, 1951) was calculated for the short-activity-day responses to measure reliability of the scales. To compare student responses to ASPECT after the two activity types, we summed the Likert-scale score (ranging from 1 to 6) of the questions within a construct (Value = 9 questions, PE = 3 questions, and IC = 4 questions), such that students could indicate Value ranging in score from 9 to 54 points, PE ranging in score from 3 to 18, and IC ranging in score from 4 to 24. We then independently modeled each construct of the survey (Value, PE, and IC) using linear mixed models. Mixed models were necessary, because we had a repeated-measures design in which the same students took ASPECT twice, once after experiencing a short-activity day and once after experiencing a long-activity day. Mixed-effects models can handle the resulting nonindependence of outcomes by including a random effect term for student (Zuur et al., 2009).

Our modeling procedure included three steps. First, we started by fitting a simple model, wherein the outcome was modeled solely as a function of the activity type (and the student random effect). Second, we fitted a complex model, in which the outcome was modeled as a function of the activity type and student demographics, including university grade point average (GPA), gender, first-generation status, and ethnicity (a categorical variable with four levels: white, Asian American, international, and underrepresented minority). After fitting the most complex model within these parameters, we selected the best-fit model by using backward selection, comparing AIC (Akaike’s information criterion) from subsequently more simple models to evaluate improvement of model fit; we considered ΔAIC < 2 to be an equivalent fit (Burnham and Anderson, 2002), in which cases we selected the model with the fewest parameters. The third step in our model selection procedure was similar to the second, but we initially fitted a full, saturated model with activity type, student demographics, and all interactions between demographics and activity type. We employed the same backward selection procedure. These models (the simple, complex, and full) test three nested, complementary hypotheses: first, that student engagement is distinguishable by activity type; second, that engagement is distinguishable by student characteristics, controlling for activity type; and third, that engagement is differentially distinguishable by student characteristics on different activity types.
Visual inspection of the residuals (Supplemental Figure S1) revealed that they were unevenly distributed, likely due to the ceiling effect in student responses (Supplemental Figure S2). A ceiling effect occurs in a survey when some respondents who gave the highest response (in our case, 6) would have responded at a higher level had they been able to do so. The ceiling effect in our data is an artifact of the Likert-scale nature of student responses (each question was answered on a 1–6 scale, and, as is typical of survey responses, students primarily answered in the upper ranges of this scale); a floor effect is in theory also possible, although our data did not display this pattern (Supplemental Figure S2). To determine whether this ceiling effect influenced our results, we fitted the final nonnull models (selected from the model selection procedure described earlier) as censored regression models (Henningsen, 2011). Censored regression models account for ceiling (and floor) effects by modeling an uncensored latent outcome in place of the censored observed outcome (Henningsen, 2011). The results from the censored regressions indicate qualitatively similar patterns (Supplemental Table S1), indicating that the results from the linear mixed models are not strongly biased.

All models were fitted in R version 3.2.3 (R Core Team, 2015). Mixed-effects models were fitted using the “lme4” package (Bates et al., 2015) and censored regression models were fitted using the “censReg” package (Henningsen, 2016). Code used for fitting models can be found in the Supplemental Material. Owing to institutional review board (IRB) restrictions, data are available only upon request.

RESULTS
Phase I: Development and Validation of Constructs Measured in ASPECT
Coding and Identification of Emergent Themes from Individual Interviews and Focus Groups. We began by recruiting students (n = 2) who had engaged in different active-learning strategies in a large introductory biology classroom for open-ended interviews (Rubin and Rubin, 2011) to answer questions centered around how they perceived the class environment. A typical 50-minute interview included a maximum of three short, intentionally broad questions; for example, What was important about today’s class? What helped your learning? Did anything make learning harder? Follow-up questions were unscripted but were consistently intended to push students to explain their reasoning as deeply as possible. The initial student-generated themes arising from these interviews focused on group dynamics, instructor language, and process-oriented features related to the activity, such as how they were directed to interact with group members.

On the basis of these initial interviews and in hopes of capturing greater depth and breadth of student experiences, we assembled a series of focus groups as described in Methods. Focus groups were progressively shifted toward questions and discussions that explored these emergent themes (group dynamics, instructor language, and process-oriented features), using a grounded theory approach (Strauss and Corbin, 1998; Glaser and Strauss, 2009). After each focus group, transcripts were coded independently by two coders (B.L.W. and L.W.-F.). Codes were iteratively revised based on frequent discussion between the coders resulting in unanimous coding at each step; the final consensus codes are shown in Table 1. Through this process, the original themes identified in the initial interviews evolved. The two themes of group dynamics and process-oriented features emerged as a single theme focused on the value of the group activity; the theme of instructor language broadened into the impact of instructor contribution on an activity; and finally, a new theme arose focused on the amount of effort students perceived themselves investing in an activity. This resulted in the three major categories listed in Table 1: 1) utility and intrinsic value of the group activity, 2) personal effort invested during the activity, and 3) instructor contribution to the activity and to student learning.

Initial Survey Item Development and Content Validity. Focusing on the themes that were most prevalent in student talk (Table 1), we developed an initial set of 26 survey items through a short series of research group writing tasks and editing sessions (Dillman et al., 2014). Content validity of the initial questions was provided through seven individual think-alouds (Gubrium and Holstein, 2002). Students read nascent survey items first silently, then aloud, and were then asked to answer the survey items out loud and to justify their reasoning for their answers. Finally, students were asked to explain or identify problematic items and to suggest alternative language if applicable. Items were then edited based on student talk during the think-alouds, with the mutual goals of maintaining coherence of student language and fidelity to the original qualitative emergent themes (Figure 2).

We next revised and refined ASPECT to conform to best practices in survey design (Dillman et al., 2014) as described in Methods and illustrated in Figure 2. The revised survey contained 20 items: eight items asking about the value students placed on the activity, seven items asking about student effort and involvement with the material during the activity, and five items asking about the instructor contribution. Three “control” questions were also included at the beginning of the survey to allow us to control for variables we hypothesized might impact student engagement: group size, prior experience with active learning, and having a friend in the group. We refer to this version of ASPECT as “20 + 3” to indicate the 20 engagement items and the three control questions.

Cognitive Testing. Next, to determine whether the language in the revised survey was easily understandable and unambiguous to students, we performed a series of cognitive testing and face validation steps. The goal of cognitive testing was to identify any confusing wording or alternative interpretations of survey items that might lead to students giving the same answer for multiple reasons (Willis, 2004). Participants (n = 6) were randomly recruited to a focus group. Each student first completed the 20 + 3 item survey in paper form, and then the entire focus group worked together to discuss possible interpretations for each item and whether the primary interpretation aligned with the intended interpretation.

Focus group participants unanimously agreed on the primary interpretation of all but one item on ASPECT. For the items agreed upon, the salient interpretations matched the goals and researchers’ intentions of the item in each case. The one potentially problematic item (One group member dominated discussion during today’s group activity) was interpreted by different members of the focus group as having
Large-Scale Face Validation of Survey Items. As an additional measure to ensure that students were interpreting the final questions as intended, we asked students in a subsequent quarter of the same course to complete the 20 + 3 item survey online (Supplemental Document S3); students were then asked to provide written explanations for why they answered the way they did for two randomly selected questions on the survey. We had a 96% response rate (n = 383), providing us with 29–40 open-ended responses per item. Student responses were independently coded by three researchers to identify the central themes emerging from student answers. Answers that were too vague to interpret or did not address the question (e.g., “I was sick that day”) were removed from analysis.

After independently coding all student responses, three researchers came together to discuss and reach consensus on whether or not students were interpreting items as intended, using an approach similar to that employed by Zimmerman and Bell (2014). Similar to the results described in Cognitive Testing,
the question regarding a dominator in the group (One group member dominated discussion during today's group activity) had multiple interpretations but was found to be consistently interpreted as relating to group equity as intended (Supplemental Document S1) and so was retained. However, as described in Phase II below, this item was removed from our final EFA analysis due to lack of correlation with other items in the survey. Only one survey item (I engaged in critical thinking during today's group activity) was identified as problematic: although the cognitive testing focus group agreed on a single meaning of this item, the larger-scale analysis of student explanations of this item (n = 29) revealed variable interpretations of the term “critical thinking.” Interpretations ranged from “the instructions were vague so I had to think critically to understand what the professor wanted” to “this activity evoked critical thinking because I had to think hard to answer the questions.” For this reason, and because there is continued debate even among experts as to the definition of critical thinking, we decided to remove this item from the subsequent analysis, resulting in a 19 + 3 item survey. One item (The instructor put a good deal of effort into my learning for today’s class) was inadvertently excluded from this large-scale validation process; however, think-alouds and cognitive testing did not reveal any conflicting interpretations of this item. A summary of themes arising from student explanations for their responses is available (Supplemental Document S1).

Phase II: Validity and Reliability of ASPECT

Refinement of Scales. ASPECT was designed to measure three constructs: 1) Value of group activity, 2) PE, and 3) IC. To determine whether survey items would be useful in measuring at least one of these constructs, we performed a pairwise correlational analysis of the 19 items remaining after face validation. Nonuseful items that consistently exhibited low interitem correlations (Spearman's rank correlation coefficient $r < 0.3$ for at least 80% of correlations) were removed (Tabachnick and Fidell, 2007). This resulted in the removal of one item (One group member dominated discussion during today's group activity) that showed no correlation with any other items in the survey, leaving 18 items.

We conducted several iterations of EFA with the remaining 18 items. There was evidence in support of both a three- and four-factor solution. The additional factor that arose in the four-factor solution contained four items, two of which cross-loaded strongly onto other factors. The four items were all related to group function (e.g., Overall, the other members of my group made valuable contributions during the activity; Group discussion during the activity contributed to my understanding of the course material). In the three-factor solution, these items combined with items intentionally constructed to capture “Value of the group activity” to form a single factor, which aligned closely with the Value theme arising from student focus groups. In our discussions with students, the value students placed on an activity was intimately connected to whether or not they perceived their group to be functioning well. Owing to the multiple instances of cross-loading in the four-factor solution and poor support for a fourth distinct construct, we chose to focus on the three-factor solution.
In the three-factor solution, two items (I knew what I was expected to accomplish; I felt comfortable with my group) loaded weakly onto multiple factors (<0.2). These two items were therefore removed from the final factor analysis, resulting in a 16-item survey. The responses to these items may be of particular interest to a researcher or instructor; thus, instead of removing them from the survey, we recommend analyzing them individually along with the third item that was not correlated with the rest of the survey items: "One group member dominated discussion during today’s group activity." Both the final 16-item survey and the complete 20 + 3 survey are available (Supplemental Documents S2 and S3).

We conducted a final iteration of the three-factor solution for the EFA with the remaining 16 items. Factor loadings for the final survey (Table 2) were consistently above the suggested minimum cutoff of 0.32 (Tabachnick and Fidell, 2007). Cronbach’s alpha, a measure of factor reliability (and therefore scale reliability), was greater than 0.78 for all three factors, providing confidence that the items within each scale are reliably measuring the same construct. Together, these findings provide evidence that the 16-item ASPECT is measuring three distinct constructs (Table 2) and that these constructs are aligned with the themes that emerged from student focus groups in phase I.

The following three factors explained 55% of the variation in student response:

- **Value of group activity:** The first factor consisted of nine items exploring students’ perception of the activity’s value for learning (e.g., Explaining the material to my group improved my understanding of it) or other reasons (e.g., I had fun during today’s group activity). Cronbach’s alpha for this scale was 0.91. This scale explained 30% of the variation in student response.

- **Personal effort:** The second factor consisted of three items that measured how much individual effort a student put into the activity (e.g., I worked hard during today’s group activity; I made a valuable contribution to my group today). Cronbach’s alpha for this scale was 0.84. This scale explained 12% of the variation in student response.

- **Instructor contribution:** The final factor included four items and measured how much effort the students perceived that the instructor put into the activity (e.g., The instructor put a good deal of effort into my learning for today’s class; The instructor’s enthusiasm made me more interested in the group activity). Cronbach’s alpha for this scale was 0.78. This scale explained 13% of the variation in student response.

**Scale Reliability.** The range of Cronbach’s alpha coefficients (Cronbach, 1951) observed for each of the three factors described above (0.78–0.91) indicates that students have a similar response pattern for the items within a given factor. To further assess the internal consistency of the scales identified in the EFA, we administered ASPECT to a similar population of introductory biology students in a consecutive quarter of the same course for which we had performed the EFA. Cronbach’s alpha coefficients for each scale ranged from 0.81 to 0.91, again providing evidence for the reliability of the scales (Supplemental Table S2). Histograms of student responses are available (Supplemental Figure S3).

**Phase III: External Validity of ASPECT**

To be a useful research tool, ASPECT must be sensitive to changing levels of student engagement with different activities. To test its ability to discriminate between activities, we compared ASPECT responses of introductory biology students during two different activity types: 1) a short-activity day with a series of 8–10 clicker-question activities centered around instructor-posed questions, and 2) a long-activity day in which students worked in groups to complete a worksheet (~30 minutes long) followed by clicker questions to check understanding. On the basis of student focus groups and our analysis of student open-ended responses to items on ASPECT, we hypothesized that students would place more value on the short activities compared with the one long activity, because students often voiced frustration regarding infrequent instructor feedback during the long activities. We also hypothesized that students would perceive the instructor putting in more effort on a short-activity day, because the instructor more frequently provides feedback to the entire class than is typical on a class day with a long activity. We did not have an a priori hypothesis about which context would be perceived to elicit more personal effort.

We first tested whether the questions on ASPECT still captured the same three constructs in this new population that had completed a day with short activities by calculating the Cronbach’s alpha for each scale (Cronbach, 1951). Because ASPECT was designed to capture student opinion about in-class activities, we reasoned that the same scales should be observed when students reflect on the short instructor-directed activities typical of a regular day, as we found when surveying students after long-activity days. This was supported by our finding that Cronbach’s alpha values for each scale on a short-activity day again fell between 0.78 and 0.91 (Supplemental Table S2).

We used a linear mixed-effect model to calculate the effect of the two different activity types on each of the three factors that make up ASPECT. We found that the ASPECT survey distinguishes between activity types, student populations, and student populations performing different activity types both in the Value students place on the activity and in the IC students perceive, but not the PE students put into the activity (Table 3). Specifically, there is evidence that activity type (Table 3, a–c) and student ethnicity (Table 3, b and c) predict a students’ Value of an activity. As we predicted, on average, students value the long activity less than the short activity (Table 3, a–c), Asian-American students value the activity more than white students (Table 3, b and c), and Asian-American students and international students both value the long activity more than white students (Table 3c). Similarly, there is evidence that activity type and student ethnicity predict a student’s perception of IC to an activity (Table 3, g–i). As we hypothesized, on average, students perceive less IC on the long activity compared with the short activity (Table 3g–i), Asian-American students perceive more IC than white students (Table 3, h and i), and international students perceive more IC on the long activity than white students (Table 3i). There is no evidence that different groups of students perceive that their PE changes in response to activity type (Table 3, d–f).
DISCUSSION

We have described here the development of ASPECT, a 16-item survey (Supplemental Document S2) that provides a rapid way to monitor students’ perception of engagement. The survey, which takes students on average 6–7 minutes to complete, provides researchers and practitioners a new tool to assess student self-reported engagement in large enrollment active-learning classrooms.

In this mixed-methods study, we triangulated qualitative analysis of students’ experience in an active-learning classroom with quantitative analysis of large-scale survey data to gain a richer understanding of student engagement. Based on the themes that emerged from qualitative student interviews and focus groups, ASPECT was intended to elicit student perception of three key constructs of cognitive and affective engagement in the active-learning classroom: 1) utility and intrinsic value of a group activity, 2) personal effort invested during an activity, and 3) instructor contribution to an activity and to student learning. EFA of student responses from a large-enrollment introductory biology class supports the assumption that ASPECT is measuring three discrete factors that align closely with these three constructs. We also provide evidence regarding the reliability of the three scales as measured by Cronbach’s alpha coefficient in a similar population under similar conditions. The internal consistency of our findings using this mixed-methods approach provides increased confidence that these three constructs are aspects of the learning experience that affect students’ engagement.

Our finding from focus groups and EFA that task value and personal effort are key factors in promoting student engagement has strong support in the sociocognitive literature (Dweck, 1986; Wigfield and Eccles, 2000; Eccles and Wigfield, 2002; Svinicki, 2004; Hidi and Renninger, 2006; Hulleman et al., 2008). Specifically, expectancy value theory predicts that perception of an activity’s value will be positively correlated with student interest and engagement (Eccles, 2005). Students place more value on activities’ value will be positively correlated with student interest and engagement (Eccles, 2005). Students place more value on tasks that they see as being either directly connected to their success, such as increasing performance on an exam or having a tangible connection to the world outside the classroom (Eccles and Wigfield, 2002; Hulleman et al., 2008). Specifically, expectancy value theory predicts that perception of an activity’s value will be positively correlated with student interest and engagement (Eccles, 2005). Students place more value on tasks that they see as being either directly connected to their success, such as increasing performance on an exam or having a tangible connection to the world outside the classroom (Eccles and Wigfield, 2002; Hulleman et al., 2008). Motivation to engage in a task is also influenced by how enjoyable the task is perceived to be and whether there is a high expectation of success in completing the task (Eccles and Wigfield, 2002; Eccles, 2005; Svinicki, 2004; Hug et al., 2005). Our qualitative work also identified the importance of instructor contribution to student engagement. Although there is a growing literature on how instructor talk may influence student participation (Myers, 2004; Seidel et al., 2015), future studies will be required to provide additional evidence that perception of instructor effort is positively correlated with student engagement.

When looking at student responses quantitatively, we detected some differences along these three constructs.

TABLE 3. The ASPECT survey is able to discriminate between types of activities (long and short) and types of students (ethnicity) on the Value and IC constructs, but PE was not predictable by student characteristics or activity type

| ASPECT construct (outcome) | Intercept | Activity type | Ethnicity | Activity type x ethnicity | ΔAIC
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Value&lt;sup&gt;1&lt;/sup&gt;</td>
<td>43.35</td>
<td>−1.09</td>
<td></td>
<td></td>
<td>4.04</td>
</tr>
<tr>
<td>b. Value&lt;sup&gt;2&lt;/sup&gt;</td>
<td>42.57</td>
<td>−1.09</td>
<td>AA</td>
<td>1.74</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Int.</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>URM</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>c. Value&lt;sup&gt;3&lt;/sup&gt;</td>
<td>43.15</td>
<td>−2.19</td>
<td>AA</td>
<td>0.68</td>
<td>22.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Int.</td>
<td>−2.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>URM</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td>d. PE&lt;sup&gt;4&lt;/sup&gt;</td>
<td>15.08</td>
<td>0.18</td>
<td></td>
<td></td>
<td>−2.53</td>
</tr>
<tr>
<td>e. PE&lt;sup&gt;5&lt;/sup&gt;</td>
<td>15.17</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>f. PE&lt;sup&gt;6&lt;/sup&gt;</td>
<td>15.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. IC&lt;sup&gt;7&lt;/sup&gt;</td>
<td>20.52</td>
<td>−1.15</td>
<td>AA</td>
<td>0.96</td>
<td>32.88</td>
</tr>
<tr>
<td>h. IC&lt;sup&gt;8&lt;/sup&gt;</td>
<td>20.07</td>
<td>−1.15</td>
<td>Int.</td>
<td>0.60</td>
<td>35.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>URM</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>i. IC&lt;sup&gt;9&lt;/sup&gt;</td>
<td>20.24</td>
<td>−1.50</td>
<td>AA</td>
<td>0.61</td>
<td>41.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Int.</td>
<td>−0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>URM</td>
<td>0.72</td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup>Table shows relationship effect sizes from linear mixed-effects models, in which students were specified as random effects. Superscripts indicate reference groups, starting models, and interpretation notes; boldface coefficients indicate significance to \( \alpha \) < 0.05. Gray cells indicate variables that were not included in the initial model; the model selection procedure is described in Methods.

<sup>2</sup>Reference level: short activity.

<sup>3</sup>Reference level: white; AA stands for Asian American; Int. stands for international; URM stands for underrepresented minority.

<sup>4</sup>Change from null model: outcome – 1 + (student random effect).

<sup>5</sup>AIC is used only to compare nested models, in this case, models modeling the same outcome.

<sup>6</sup>Simple model was specified as Outcome – Treatment + (student random effect).

<sup>7</sup>Complex model was specified as Outcome – Treatment + Demographics + (student random effect). Student demographics included university GPA, ethnicity, first-generation status, and gender.

<sup>8</sup>Full model was specified as Outcome – Treatment + Demographics + Treatment x Demographics + (student random effect). Student demographics included university GPA, ethnicity, first-generation status, and gender.
Specifically, ASPECT distinguished between activity types in terms of how much value students place on the different activities and how much instructor contribution they perceive but not the amount of personal effort they put forth. Unlike the direct association of motivation with task value (Eccles and Wigfield, 2002), the PE factor measured by ASPECT does not appear to be correlated with task value.

Limitations of ASPECT

ASPECT is not intended as a psychometric analysis of the mental construct of student engagement, for which additional validation beyond the scope of our observations would be necessary. Instead, our interest was in developing a way to systematically collect students' self-reported involvement during in-class activities. Student perception of engagement is just one measure of engagement and, as with all self-report data, contains inherent biases. To gain a more holistic view of student engagement, one could administer ASPECT in conjunction with other tools designed to measure specific aspects of student behavior such as motivation or sense of belonging (Ryan et al., 1983; Pintrich et al., 1993; Hagerty and Patusky, 1995).

This survey was developed and validated with students in a single introductory biology course at one university and may therefore not be applicable to other course levels in biology, other science, technology, engineering, and mathematics (STEM) majors, or other populations. However, we found ASPECT reliably measures the same constructs in different iterations of the same course taught by different instructors. Furthermore, we have no reason to suspect that ASPECT would be differentially effective in different course levels or in other populations, although further field-testing is necessary to confirm this assumption.

Our intention is for ASPECT to be used to compare different active-learning strategies. Here, we show that ASPECT can discriminate between two different types of active-learning exercises that varied with respect to length and instructor feedback. Further field-testing of the instrument will be important to assess whether ASPECT can differentiate between active-learning exercises that differ in other elements. The best use of this survey will require validation in new classroom environments to ensure that the language and interpretation of ASPECT questions are meaningful for the new population (Lave and Wenger, 1991).

Implications for Research and Teaching

Research. As we elaborate below, ASPECT can be useful for researchers. First, as a research tool that targets student engagement, ASPECT may give insight into the “leaky STEM pipeline”; furthermore, student populations are differentially affected by active-learning activities, and ASPECT could lend insight into the specific differences; and finally, active learning as a research field is moving toward a finer-grained understanding of the most effective aspects of active learning, and one element of efficacy is student engagement.

Defined as the leaky STEM pipeline, only half of the students who enter college in the United States intending to major in a STEM discipline end up completing a science degree (President’s Council of Advisors on Science and Technology, 2012). This indicates that far too many talented and interested students are lost along the way (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2007; Drew, 2011; Dasgupta and Stout, 2014). Students who exit STEM come disproportionately from backgrounds historically underrepresented in STEM and report social threats and unwelcoming atmosphere as major factors behind their decisions to leave STEM education (Steele, 1997; Seymour and Hewitt, 1998; London et al., 2012; Graham et al., 2013). The recent shift toward more student-centered learning in STEM classrooms is expected to increase retention (Kvam, 2000; McConnell et al., 2003; Haak et al., 2011), as active learning has the potential to disproportionately benefit underrepresented groups’ learning outcomes (Springer et al., 1999; Haak et al., 2011; Eddy and Hogan, 2014). However, an important element to consider in retention of STEM majors is how students engage in classrooms. If paired with retention data, ASPECT could identify areas in which students’ value, personal effort, and perceived instructor effort are correlated with attrition.

Furthermore, different student populations are disproportionally affected by active-learning activities (Springer et al., 1999; Haak et al., 2011; Eddy and Hogan, 2014). Underlying cultural factors, including gender spectra, race/ethnicity, and/or socioeconomic backgrounds are thus likely to impact student engagement during active learning. Our results suggest that, not surprisingly, students from different demographic groups perceive the same in-class activity differently. Specifically, Asian-American students saw more value in the group activity than white students. ASPECT could also detect differences in how demographic groups valued different types of activities. For example, the longer activity was more favorably perceived by both international students and Asian students compared with white students. Interestingly, student GPA did not predict whether or not students placed more value on the group activity, suggesting that, all else being equal, students in the top and bottom of the class do not have different perceptions of the activities.

Finally, as the biology education research field moves toward a finer-grained analysis of what makes active learning impactful, considering student engagement via ASPECT may prove beneficial. Because the modes and implementation of active learning vary widely (Andrews et al., 2011; Freeman et al., 2014), it will be important to continue to monitor engagement along with learning outcomes of different demographic populations to determine whether all students are engaging equally. Comparing student responses on ASPECT could enable researchers to assess the impact of different elements of an activity and could provide insight into why one activity is more beneficial than another. For example, by comparing two activities that differ in only one element, researchers could identify pedagogical approaches that influence how much value and personal effort student place on an activity. Additionally, one could compare activities that vary in either their mode of student interaction (e.g., with or without designated group roles) or the method of group assignment (e.g., self-selected, randomly assigned, or instructor assigned). In this way, one could determine, for example, whether students place more value and perceive themselves putting more personal effort into activities with highly structured roles for each individual. Pairing ASPECT data with follow-up student interviews or focus groups could help us to ascertain why students place more value on a particular
activity. Finally, as discussed earlier, ASPECT could help gain perspective on how students’ unique characteristics, such as ethnicity, influence their experiences during an active-learning exercise.

Teaching. In addition to being useful to researchers, ASPECT can be helpful for practitioners. Teachers are faced with a myriad of decisions to make daily; when implemented strategically, ASPECT can be one avenue for data-driven decision making. As we enumerate below, by determining levels of student engagement with ASPECT, instructors can inform their decision to continue, modify, or discontinue an activity; and to inform their own teaching practices, instructors can use ASPECT for comparison between activities, student populations, and, potentially, between instructors.

First, as ASPECT is a measure of the level of student engagement on a particular activity, instructors can use these data to inform their decisions to revise or discontinue activities. There are many different and effective active-learning strategies (Tanner, 2013) and a number of different ways to implement active learning (Borrego et al., 2013). After implementing a strategy or teaching an activity, instructors need to decide whether that activity was effective and whether they want to use it again, modify it, or avoid it in the future. ASPECT data can help inform this decision by providing data on the level of student engagement.

Additionally, instructors can use ASPECT data to compare activities, student populations, or, potentially, instructors. Comparing results on ASPECT between two activities can help inform prioritizing one activity over another. Furthermore, as previously discussed, ASPECT can distinguish between student populations, so if an instructor knows that a particular demographic group in a class is struggling academically, they may be able to employ ASPECT to determine whether engagement in this struggling population is also stunted. Finally, ASPECT can be used to determine whether there are certain aspects of instructor behavior that are most effective at engaging students. For example, comparing two activities with different instructor framing techniques could inform how to best set up an activity for students. In this way, ASPECT may also serve as a reminder for novice active-learning instructors that there are many different elements to consider when implementing student-centered strategies, including fostering functional groups and helping students see the value in an activity. Similarly, in a mentoring or coaching relationship, instructors might also be able to compare student engagement between two instructors; paired with additional classroom observation data about instructor habits (for example from a tool like PORTAAL [Eddy et al., 2015] or COPUS [Smith et al., 2013]), instructors may be able to enhance their own soft skills to increase engagement in their own classrooms.

CONCLUSIONS

Our goal with this work was to develop a survey to systematically gather student perception data to compare relative student engagement levels across various active-learning strategies. ASPECT differs from other instruments that have been designed to measure student experiences during active learning (Visschers-Pleijers et al., 2005; Pazos et al., 2010) in that we intentionally designed this survey to be widely applicable for different types of active learning. Our findings suggest that classroom culture, including small-group dynamics and instructor enthusiasm, could influence students’ willingness and inspiration to engage in difficult STEM learning tasks. Gathering more information through tools such as ASPECT will help us better understand potential barriers presented by an active-learning environment (Malcom and Feder, 2016) and ideally develop strategies that increase engagement of all students. Our hope is that ASPECT will provide researchers and instructors alike with a tool to rapidly evaluate active-learning strategies from the perspective of the learner. These data can then be used, in conjunction with student performance data, focus group data, and even classroom observation data, to help inform instructional choices in the classroom.

ACKNOWLEDGMENTS

We thank M. P. Wenderoth and J. Doherty for their insightful and detailed comments on earlier versions of this article and the University of Washington Biology Education Research Group as a whole for valuable discussions on this topic. We also thank Kayla Evans and Dr. Linda Martin-Morris for sharing their analysis of student behaviors in the classroom. This research was completed under approved IRB protocol #44438 at the University of Washington and was supported in part by an award from the National Science Foundation (NSF DUE 1244847).

REFERENCES

SUPPLEMENTAL MATERIALS

Figure S1: Residual plots from final models
Figure S2: Ceiling Effect
Figure S3: Histogram of student responses
Table S1: Censored regression model results
Table S2: Reliability and external validity of ASPECT: Cronbach’s alpha coefficients
Document S1: Themes of student responses
Document S2: ASPECT final survey (16 items + 3 control questions)
Document S3: ASPECT long version of survey administered online (20 items + 3 control questions)
Figure S1: Residual plots from final models indicate that the residuals are not evenly distributed, rather, there is a strong ceiling effect.
Figure S2: The ceiling effect in student responses is evident in the raw data – students tend to answer all questions favorably thus “max out” the scale of possible responses. In theory there could also be a floor effect wherein students are “not allowed” to answer below a certain value. However, student responses do not approach the floor thus the floor effect is not likely to influence inference. All levels shown are bounded between possible student responses.
Figure S3: Histograms of student responses on all ASPECT items. These histograms were used to assess if responses to any items were problematically narrow and lacking in distribution. Each set of diagrams corresponds to a single factor or to those three questions that did not load strongly onto any of the three factors.

Items that loaded onto the Personal Effort factor:

- I worked hard during today’s group activity
- I made a valuable contribution to my group today
- I was focused during today’s group activity

Items that loaded onto the Value of Group Activity factor:

- Explaining the material to my group improved my understanding of it
- Group discussion during the activity contributed to my understanding of the course material
- Overall, the other members of my group made valuable contributions during the group activity
- I had fun during today’s group activity
- The group activity stimulated my interest in the course material
- I would prefer to take a class that includes this PCR group activity over one that does not include this PCR group activity
- The group activity increased my understanding of the course material
Items that loaded onto the Instructor Contribution factor:

- The instructor's enthusiasm made me more interested in the group activity
- The instructor seemed prepared for the group activity
- The instructor put a good deal of effort into my learning for today's class
- The instructor and TAs were available to answer questions during the group activity

Other three items that did not load well onto any of the three factors:

- One group member dominated discussion during today's group activity
- I felt comfortable with my group
- I knew what I was expected to accomplish during the group activity
Table S1: Results from the censored regression models are qualitatively similar to the results from the linear mixed models thus either model type is appropriate for inference. The censored regression, however accounts for the ceiling effect thus are likely to be more accurate for precise estimation. Table shows relationship effect sizes from censored regression models, where students were specified as random effects. Grey cells indicate variables that were not included in the initial model; model selection procedure is described in Methods. Superscripts indicate reference groups and starting model, boldface coefficients indicate significance to \( \alpha < 0.05 \).

<table>
<thead>
<tr>
<th>ASPECT Construct (outcome)</th>
<th>Intercept</th>
<th>Activity Type(^1)</th>
<th>Ethnicity(^2)</th>
<th>Activity Type x Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value(^3)</td>
<td>43.60</td>
<td>-1.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value(^4)</td>
<td>42.64</td>
<td>-1.05</td>
<td>AA 2.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Int. 1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>URM 1.02</td>
<td></td>
</tr>
<tr>
<td>Value(^5)</td>
<td>43.26</td>
<td>-2.26</td>
<td>AA 0.88</td>
<td>AA:Long 2.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Int. -1.86</td>
<td>Int.:Long 5.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>URM 1.51</td>
<td>URM:Long -0.96</td>
</tr>
<tr>
<td>Personal Effort(^3)</td>
<td>15.37</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal Effort(^4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal Effort(^5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Contribution(^3)</td>
<td>21.05</td>
<td>-1.33</td>
<td>AA 1.28</td>
<td></td>
</tr>
<tr>
<td>Instructor Contribution(^4)</td>
<td>20.44</td>
<td>-1.33</td>
<td>Int. 0.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>URM 0.45</td>
<td></td>
</tr>
<tr>
<td>Instructor Contribution(^5)</td>
<td>20.60</td>
<td>-1.64</td>
<td>AA 0.94</td>
<td>AA:Long 0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Int. -0.10</td>
<td>Int.:Long 1.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>URM 0.86</td>
<td>URM:Long -0.79</td>
</tr>
</tbody>
</table>

\(^1\)Reference level: short activity.  
\(^2\)Reference level: white; AA stands for Asian American, Int. stands for International, URM stands for Under Represented Minority.  
\(^3\)Simple model was specified as Outcome ~ Treatment + (student random effect).  
\(^4\)Complex model was specified as Outcome ~ Treatment + Demographics + (student random effect). Student demographics included university GPA, Ethnicity, First Generation Status, and Gender.  
\(^5\)Full model was specified as Outcome ~ Treatment + Demographics + Treatment*Demographics + (student random effect). Student demographics included university GPA, Ethnicity, First Generation Status, and Gender.
**Table S2**: Comparison of Cronbach’s alpha values resulting from exploratory factor analysis of student responses to ASPECT administered to different populations

<table>
<thead>
<tr>
<th>Factor</th>
<th>Fall 2014&lt;sup&gt;a&lt;/sup&gt; Long Activity Day&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Winter 2015 Long Activity Day</th>
<th>Winter 2015 Short Activity Day&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of Group Activity</td>
<td>0.91&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Personal Effort</td>
<td>0.84</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>Instructor Contribution</td>
<td>0.78</td>
<td>0.84</td>
<td>0.82</td>
</tr>
</tbody>
</table>

<sup>a</sup>Quarter of introductory biology course to which the survey was administered  
<sup>b</sup>Cronbach’s alpha values  
<sup>c</sup>Long activity day refers to a class day in which students completed a ~30 minute worksheet followed by instructor feedback  
<sup>d</sup>Short activity day refers to a class day in which students completed 8-10 “clicker-question” activities centered around instructor-posed clicker questions
Document S1: Themes in large-scale validation of responses to ASPECT items

Factor 1: Activity’s influence on learning

Explaining the material to my group improved my understanding of it:  
N= 37
Themes:

• Dominant theme: explaining allows them to check their understanding (‘I think by explaining the material, not only you are able to have a better understanding of it but your group members can also help point out if what you are saying is right or wrong’)
• Some vague answers indicating that students may not be referencing the activity specifically but be considering the general value they see in explaining (‘I'm am the kind of learner that likes to talk out my thoughts in order to solidify them. Thus explaining concepts to my group helped fortify them in my head’)

Having the material explained to me by my group members improved my understanding of the material:  
N = 40
Themes:

• Deciding whether or not listening helped them learn (‘the discussion helped clear up questions or confusions I had about the processes and they could explain it in ways aside from the text that was understandable’ or ‘in certain cases as a group member explained something to me, they did not know exactly what they were talking about or were completely wrong’)
  o This was influenced primarily by
    ▪ Whether or not they trusted their groupmates
    ▪ Whether or not they were comfortable with groupmates

Group discussion during the [topic] activity contributed to my understanding of the course material:  
N = 38
Themes:

• Improved understanding because discussing topic (‘it's easier to understand a concept and get a full picture of it when you discuss it with others’)
  o This was primarily because:
    ▪ Multiple perspectives heard
    ▪ Someone in group was knowledgeable
• Over half referenced how peer’s contribution influenced their understanding

I had fun during today's [topic] activity:  
N = 33
Themes:

• Many students referenced fun as enjoying learning (‘the [activity was] fun because the group can work at their own pace and there is an interactive element involved’)
• A few talked about fun like it was enjoyable (‘I completed it with a friend and it was fun’)
  o Having a friend in the group or not influenced fun (‘I really have no social connections to anyone in the class and I tend to feel like I'm bothering a group of friends who would rather not have me latched on when we're expected to work in groups’)
  o Lack of confidence in ability to learn from activity limited fun (‘my confidence in the material was not strong, therefore, it made the [activity] more frustrating than fun’)


Overall, the other members of my group made valuable contributions during the [topic] activity.
N = 38
Themes:

• Need knowledge to valuably contribute to group (‘members in my group had read and were able to contribute helpful information’)
• Participation is considered a valuable contribution (‘my other group member was actively participating and helping to answer the questions just as much as I was’)
• 2-3 people said hearing different perspectives from group members was a valuable contribution (‘discussing the questions with group members was helpful since everyone in the group had a different way of thinking about the material’)
• Value was always in reference to the participant not value to the group

I would prefer to take a class that includes this [topic] activity over one that does not include this activity:
N = 29
Themes:

• Frequently referenced this type of long in-class activity in general (‘I don't learn very well through the [long class] activities. I'd much rather just listen to a lecture and take notes’ or ‘having the visual aids on the paper to refer to really helps me solidify my understanding of what is going on’)
• Gave reasons why they did or did not like the long in-class activities (many different explanations)
• Activity features limited preference (‘today's activity was entirely too long unless you somehow knew every answer right away’)

I am confident in my understanding of the material presented during today’s [topic] activity.
N = 35
Themes:

• Explained how well they did or did not understand the material for the day (‘I actually sometimes, am more confused if the exact correct answer is not presented straight away’)
• Confidence limited by lack of preparation on their part or procedural problems (‘I was not fully prepared when coming into class today. However, having a group discussion helped clarify any confusion on the matter’)

The [topic] activity increased my understanding of the course material:
N = 49
Themes:

• Described how activity helped them learn (‘the group activity was helpful because it allowed me to discuss concepts with my classmates and get a chance to teach as well. It also helped me figure out what parts of the material I was not confident in’)
• Procedural problems (like bad wording on questions) could limit learning
• Many students compared to their preference for lecture (‘I personally like getting information through a lecture style and then applying the information in a group setting’)

The [topic] activity stimulated my interest in the course material:
N = 33
Themes:

- How activity stimulated interest: primarily working with others was what engaged them with material (‘talking with the group engaged me with the subject, thus making it more interesting’)
- Lack of understanding limited ability to be stimulated

Factor 2: Personal Effort

I made a valuable contribution to my group today:
N = 33
Themes:

- Valuable contribution is giving correct information (‘I helped explain concepts regarding [topic] to my partner who did not know about these topics’)
- Small number said that their contribution was limited by their confusion (‘once I started to understand the topic, I was able to contribute to the group and help them understand’)

I was focused during today's [topic] activity:
N = 39
Themes:

- Personally engaged (‘I was engaged in discussion with my group members about the [activity]’)
  - Primarily influenced by:
    - Structure of activity (‘if I had my own worksheet I think I would be able to pay attention more’)
    - Difficulty level and how interesting activity was influenced focus (‘[topic] is still a little 'iffy' to wrap my head around, so today I made it my goal to stay focused and work on improving my understanding of the topic’)
    - Relevance of activity (‘I was focused during today’s group activity since this will make me understand the PCR concepts better and this will help for studying my final’)
    - Outside of class events (like a holiday) (‘being the Monday after break I can't admit I was at 100% focus today’)

I worked hard during today's [topic] activity:
N = 39
Themes:

- Explained whether or not worked hard with:
  - Whether or not contributed to group (‘I felt like I was an active participant during the group activity today’)
  - Whether or not trying to learn/interested (‘I really enjoyed this [activity], and because of that, I worked hard’)
  - Whether or not their group functioned well (‘it was frustrating today because I really wanted to work on the worksheet but nobody seemed interested’)
  - Staying focused (‘I found it easy to be distracted given the setting in the large lecture hall’)

Factor 3: Instructor Contribution

The instructor's enthusiasm made me more interested in the [topic] activity:
N = 35
Themes:

- Instructor’s enthusiasm made students value activity (*our instructor* is engaging with her teaching style and this [activity] in particular was helpful…)
- About half referenced instructor’s effort generally (*our instructor* makes all her classes enjoyable because you can tell she really enjoys the subject)

The instructor put a good deal of effort into my learning for today’s class:  
N = 0  
Note: this item inadvertently did not go through the large-scale face validation step.

The instructor seemed prepared for the [topic] activity:  
N = 45  
Themes:

- Prepared seemed to be determined:
  - Primarily by instructors ability to respond to questions in class (*She never seemed confused and could explain each aspect of the topic in depth*)
  - Secondly by prepared with materials ready (*our instructor* had worksheets ready for everyone, and had clicker questions to test our progress and understanding)
- ~1/3 gave general responses (*our instructor* is always prepared)

The instructor and TAs were available to answer questions during the [topic] activity:  
N = 42  
Theme:

- Mostly referencing if instructor/TA available when needed help (*when i raised my hand to ask a question, either a TA or the instructor immediately came to help*)

Items below were removed after EFA analysis due to poor inter-item correlation or weak/no loading on factors

One group member dominated discussion during today’s [topic] activity.  
N = 34  
Themes:

- group equity (*everyone contributed to the discussion*)
  - Influenced primarily by
    - personal approach (*I tend to dominate discussions a bit*)
    - valuing each other’s contributions (*there were people that talked a little bit more but we all helped each other answer the questions*)
    - structure of activity (1 sheet per group)(*The group member with the paper did most of the reading and writing and would therefore begin (sic) discussion by giving her thoughts*)
    - having a friend in the group (*we were all friends and valued each others answers*)

I knew what I was expected to accomplish during the [topic] activity  
N = 33  
Themes:
• clarity of questions and instructions on activity (‘the directions are pretty clear on the worksheet’)
• clarity of instructor’s verbal instructions/feedback (‘[our instructor] kept us updated about which questions we would be expected to complete in a certain time frame’)
• relevance of activity for their learning (‘I knew that [the activity] would help me have a better understanding about [topic x] for the test’)

I felt comfortable with my group.
N= 37
Themes:
• Ability to share ideas freely (‘they would not care if I said a wrong answer’)
• Familiarity with group members (‘I knew my group members’)
• Group members were all engaged/interested (‘some members would rush others with the goal to finish it rather than stop, ask questions, and discuss’)
  o influenced primarily by:
    ▪ whether or not had a friend in the group (‘all of my group members were my friends’)
    ▪ general friendliness of groupmates (‘everybody was nice’)
    ▪ size of group (‘I get to ask questions in a smaller group instead of the whole class’)
    ▪ personal attributes (‘I am typically a pretty outgoing and easy to talk to person’ or ‘I'm scared of people, human interaction, and just talking in general’)

I engaged in critical thinking during today’s [topic] activity*:
N = 29
Themes:
• Students had a wide variety of definitions of critical thinking including:
  o applying prior knowledge (‘I thought back to different course materials in order to fully answer the activity questions’)
  o teaching themselves (‘we were able to use smaller questions to guide a thought process that allowed us to come to the solutions ourselves’)
  o navigating a challenging activity (‘a lot of it was confusing, so that's why I said I engaged in critical thinking’)
  o overcoming external problems (‘it's harder for me to focus when it's louder in [the lecture] hall’)

*This item was removed from the final ASPECT due to ambiguous interpretations of the term “critical thinking”
Note: In our administration of the survey, we inserted the name of the specific activity students completed in that day's class (e.g. eukaryotic gene regulation activity) in order to remind students of the activity to which the survey is referring. It may be possible to generalize the survey to refer to “group activity”.

Instructions for students:
All questions in this survey refer to today’s class in which you completed an activity on [insert topic name]. Your responses on this survey will be used to evaluate how we teach this topic in future Biology classes. Your instructor will not know whether you completed this survey or how you answered the questions, but your effort will impact the experience of future students in this series.

Setup questions – to be used to control for student group experience
A) During class today, you and your classmates completed a [insert topic name] activity in a group. How many students (including you) worked in your group?
   Possible Answers: 1 (just me)
   2
   3
   4
   More than 4

B) Are you friends with at least one person that was in your group?
   Possible Answers: Yes
   No

C) As a college student, have you been asked to work with other students during class time in large lecture courses (over 100 students)? Do not include this class in your answer.
   Possible Answers: Yes
   No, other large lecture courses I have taken have not asked me to do this
   No, this is the first large lecture course I have taken

For questions 1-16, students answered on a 6-point Likert Scale from Strongly Disagree to Strongly Agree.

Instructions for students
The following questions ask you about your experience with the [insert topic name] activity that you completed today. Please rate how strongly you agree or disagree with each of the following statements.

1) Explaining the material to my group improved my understanding of it*
2) The instructor's enthusiasm made me more interested in the [insert topic name] activity
3) Having the material explained to me by my group members improved my understanding of the material**
4) Group discussion during the [insert topic name] activity contributed to my understanding of the course material
5) The instructor put a good deal of effort into my learning for today's class.
6) I had fun during today's [insert topic name] activity.
7) Overall, the other members of my group made valuable contributions during the [insert topic name] activity.
8) The instructor seemed prepared for the [insert topic name] activity.
9) I would prefer to take a class that includes this PCR group activity over one that does not include this [insert topic name] group activity.
10) I am confident in my understanding of the material presented during today’s [insert topic name] activity.
11) I made a valuable contribution to my group today.
12) The instructor and TAs were available to answer questions during the [insert topic name] activity.
13) The [insert topic name] activity increased my understanding of the course material.
14) I was focused during today's [insert topic name] activity.
15) The [insert topic name] activity stimulated my interest in the course material.
16) I worked hard during today's [insert topic name] activity.

* Included an extra answer choice: “I did not explain material to my group today”
** Included an extra answer choice: “A group member did not explain the material to me”
Document S3: ASPECT (long version piloted in our study)

For questions 1-20, students answered on a 6-point Likert Scale from Strongly Disagree to Strongly Agree.

Instructions for students
The following questions ask you about your experience with the [topic name] activity that you completed today. Please rate how strongly you agree or disagree with each of the following statements.

1) Explaining the material to my group improved my understanding of it. *
2) The instructor's enthusiasm made me more interested in the group activity.
3) Having the material explained to me by my group members improved my understanding of the material. **
4) Group discussion during the activity contributed to my understanding of the course material.
5) The instructor put a good deal of effort into my learning for today's class.
6) I had fun during today's group activity.
7) Overall, the other members of my group made valuable contributions during the group activity.
8) The instructor seemed prepared for the group activity.
9) I would prefer to take a class that includes this PCR group activity over one that does not include this [topic] activity.
10) I am confident in my understanding of the material presented during today’s group activity.
11) I made a valuable contribution to my group today.
12) The instructor and TAs were available to answer questions during the group activity.
13) The group activity increased my understanding of the course material.
14) I was focused during today's group activity.
15) The group activity stimulated my interest in the course material.
16) I worked hard during today's group activity.
17) I felt comfortable with my group.***
18) I knew what I was expected to accomplish during the group activity.***
19) One group member dominated discussion during today’s group activity.***
20) I engaged in critical thinking during today’s [topic] activity.****

* Included an extra answer choice: “I did not explain material to my group today”
** Included an extra answer choice: “A group member did not explain the material to me”
*** Removed from final EFA due to weak loading onto factors
**** Removed from final version due to ambiguous wording