122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Making Value for Society
Paper ID #13261

# Measuring Student Perceptions of Engineering Classroom Activities and the Use of Such Measures by STEM Faculty: The Development of the Student Class Activity and Engagement Instrument

**Dr. David L. Little II, Oregon State University**

Dr. Little is a post-doc scholar at Oregon State University and a graduate of the STEM Education program at the University of Kentucky. He specializes in education measurement across the STEM disciplines, sustainability education at the post-secondary level, and interdisciplinary research and teaching within the STEM disciplines.

**Dr. Kathleen Quardokus Fisher, Oregon State University**

Dr. Kathleen Quardokus Fisher is a post doctoral scholar at Oregon State University. She is currently participating in a project that supports the use of evidence-based instructional practices in undergraduate STEM courses through developing communities of practice. Her research interests focus on understanding how organizational change occurs in higher education with respect to teaching and learning in STEM courses.

**Dr. Shane A. Brown P.E., Oregon State University**

Shane Brown is an associate professor in the School of Civil and Environmental Engineering at Oregon State University. His research interests include conceptual change and situated cognition. He received the NSF CAREER award in 2010 and is working on a study to characterize practicing engineers' understandings of core engineering concepts.

**Dr. Milo Koretsky, Oregon State University**

Milo Koretsky is a Professor of Chemical Engineering at Oregon State University. He received his B.S. and M.S. degrees from UC San Diego and his Ph.D. from UC Berkeley, all in Chemical Engineering. He currently has research activity in areas related engineering education and is interested in integrating technology into effective educational practices and in promoting the use of higher-level cognitive skills in engineering problem solving. His research interests particularly focus on what prevents students from being able to integrate and extend the knowledge developed in specific courses in the core curriculum to the more complex, authentic problems and projects they face as professionals. Dr. Koretsky is one of the founding members of the Center for Lifelong STEM Education Research at OSU.

**Dr. Jana Bouwma-Gearhart, Oregon State University**

Jana L. Bouwma-Gearhart is an associate professor of STEM education at Oregon State University. Her research widely concerns improving education at research universities. Her earlier research explored enhancements to faculty motivation to improve undergraduate education. Her more recent research concerns organizational change towards postsecondary STEM education improvement at research universities, including the interactions of levers (people, organizations, policy, initiatives) of change and documenting the good, hard work required across disciplinary boundaries to achieve meaningful change in STEM education.

# Measuring Student Perceptions of Engineering Classroom Activities and the Use of Such Measures by STEM Faculty: The Development of the Student Class Activity and Engagement Instrument

Introduction

Effective classroom practices that interactively engage students have been directly linked to higher learning gains [1, 2, 3, 4, 5, 6, 7]. Yet, outside of time-consuming observation protocols [8, 9], little systematic work has been done in characterizing classroom activities, due in part to a lack of proper instrumentation [10]. Additionally, students can engage in activities in very different ways than instructors intend [1, 2], so it is important to account for the student perspective. This paper reports current progress towards an effort to develop and rigorously assess a student survey instrument aimed at this purpose in post-secondary STEM classrooms: the Student Class Activity and Engagement Instrument (SCAEI).

The SCAEI is based on Chi's Interactive, Constructive, Active, and Passive (ICAP) framework [1, 2]. The ICAP framework differentiates classroom activities into four categories (interactive, constructive, active, or passive) and posits a hierarchy of which types of classroom activities afford greater student learning outcomes (interactive > constructive > active > passive). Table 1 provides definition of the categories. Consider, for example, possible activities of a student when working on a group assignment. If a student chooses to sit and listen and does nothing else during the group assignment, the student's activity would be categorized as passive. If the student chooses to work with partners and those interactions lead to a deeper understanding of the material than she/he could develop individually, this activity would be classified as interactive. Or, each student might each individually answer a part, but fail to discuss his/her part with the others, thereby bypassing the interactive component of the group assignment. This classroom activity would be classified as either constructive (if he/she synthesized different sources of course content to generate the individual responses) or active (if he/she recalled information or algorithmically performed a known problem solving technique). The category "active" is specific to ICAP should be distinguished from the broader term "active learning" [11].

**Table 1**
*Definitions and Guiding Examples for ICAP Dimensions Used in SCAEI Development*

| |
|---|
| Interactive - Activities in which the student participates in an activity with another person (typically either the instructor or a classmate) where the resulting learning or cognitive engagement was not possible without another person's presence or input [1, 2]. Examples include working with peers to construct a deeper understanding of course material through group activity or interacting with the instructor in a way that augments understanding [1, 2]. |
| Constructive - Activities in which the cognitive load of students is heightened, and asks them to "produce outputs that contain ideas that go beyond the presented information" [1, p. 77]. Examples include creating diagrams to organize course content, rephrasing the instructors lecture into the student's own words, etc. [1, 2] |
| Active - Activities in which students are only cognitively engaged at a basic level, such as note taking, gesturing or pointing, or repeating information [1, p. 77]. Examples include writing or highlighting notes, watching a video and pausing/playing/rewinding/fast-forwarding, looking at a |

| |
|---|
| manipulative/physical representation, etc. [2] |
| Passive - Activities in which the student is minimally engaged, such as just sitting and watching/listening to a lecture or video without doing anything else. Examples include, listening without connecting what is heard to prior course content, etc. [1, 2]. |

The ICAP framework was used as the theoretical foundation for the SCAEI for three major reasons. First, the ICAP framework "compares and contrasts one learning activity with another learning activity during a learning phase, rather than comparing a student's activity and an instructor's activity" [1, p. 75]. Since student activity and instructor actions can be orthogonal [2], measuring frequencies of specific student activities can provide additional information about the classroom environment than can be garnered through other sources (i.e., observation protocols). Second, the ICAP framework offers a way to classify and categorize classroom activity along clearly defined dimensions; this is a requirement for any framework being operationalized for measurement [12]. Also, Chi argues that different learning activities "can be assessed in terms of their frequency of occurrences and can be coded in a variety of ways and analyzed as evidence of mediators of learning" [1, p. 76]. This offers a clear technique for measurement (i.e., frequencies) that could then be used to describe the classroom environment. Third, while interactive classroom activities have been shown to be most effective in increasing student learning outcomes [1, 2], non-interactive types of classroom activities are still of importance across the STEM disciplines; the ICAP framework provides a way to measure classroom activities from all of these perspectives [1, 2].

## Purpose and Significance of the SCAEI

The purpose of the SCAEI instrument is to measure post-secondary STEM students' self-reports of classroom activities as delineated by the ICAP framework. The SCAEI is not intended to be used to evaluate teaching effectiveness or student learning, but rather to measure how students perceive they are participating in classroom activities (with respect to the categories outlined in the ICAP framework). These data can be used (1) by instructors to improve practice, (2) by researchers to relate classroom environment to learning outcomes, and (3) by administrators to make strategic decisions about curricular delivery and support structures. In short, the SCAEI provides a tool to contribute to our developing understanding of how classroom activities and environments support and foster learning through describing how students cognitively engage in them.

## Constructing the SCAEI

To ensure the validity of the SCAEI, the ICAP framework was used to guide the construction for the SCAEI. A team of three item writers worked together to develop items for the SCAEI (see Appendix A for the instrument), where items were developed to target a unique dimension of the ICAP framework (i.e., items designed to measure only interactive activities, etc.). Items were developed this way to ensure content validity (i.e., what is being measured is measured appropriately and comprehensively) and construct validity (i.e., what is being measured is beneficial and informative to measure and the measure functions in predictable ways). As a validity check, each item was assessed by content experts across the STEM disciplines to ensure (1) a common, unique dimension of the ICAP framework could be assigned for each item (more details on this in the methodology section) across the STEM disciplines and (2) that learning activities common across the STEM disciplines were represented in the items. These parts of instrument development are sometimes ignored [12], as researchers:

> often throw together or dredge up items and assume they constitute a suitable scale. These researchers may give no thought to whether the items share a common cause (thus constituting a scale), share a common consequence (thus constituting an index), or merely are examples of a shared superordinate category that does not imply either a common causal antecedent or consequence (thus constituting an emergent variable) [12, p. 13].

Completing these steps ensures that developed instruments have predictable outcomes that can be used to assess instrument and item functioning, which serve to assess the validity of an instrument [12].

As items were developed, instrument specifications were also decided, as recommended [12]. These instrument specifications included measurement method and response format. Measurement method refers to how an instrument will be scored (e.g., Thurstone scaling [12], Guttman scaling [12], scales with equally weighted items, scales scored using item response theory methods). Item response theory (IRT) methods were chosen for the SCAEI since they can detect item bias towards subpopulations (i.e., differential item functioning; see [13]), can assess model-fit at the item level [13], and can be used to create “short-versions” of instruments that produce similar measures as the original instrument with the same degree of reliability [13]. Response format refers to how respondents will respond to each item (e.g., Likert scales, binary scales, semantic differential scales, and visual analog). A polytomous Likert scale was chosen for the SCAEI since polytomous response scales are more accurate than binary response scales [14] and because it is a response format that is likely familiar to members of the target population (STEM undergraduate students). Moreover, a unipolar response scale was chosen for the SCAEI because unipolar response scales are recommended for instruments measuring frequency [15]. The response scale chosen for the SCAEI was “never”, "rarely in a term", "monthly", "weekly", and "every class."

## Challenges of Using the ICAP Framework in Constructing the SCAEI

Two main difficulties were faced in item writing for the ICAP framework. These difficulties are shared to further elucidate the instrument development process for others wishing to develop instruments to assess classroom phenomena.

One concern was informed by Chi and Wylie's [2] comment that outputs associated with constructive activities can either be externalized or internalized. However, internalized outputs pose problems for researchers investigating student engagement because they are unobservable [2]. That is to say, if a student is just "sitting and listening" to lecture, one must be careful to not simply label this activity as passive due to the possibility that the student may be internalizing constructive behavior such as mentally connecting what is heard to prior material and thus constructing new knowledge [2]. Chi and Wylie [2] offer further that:

> One of the most difficult activities to classify is problem solving, because it depends on how a specific student is doing it and the context in which it was taught. For example, solving an algebra problem can be classified as an active behavior when the student matches the conditions of application of an equation and merely applies the formula of an equation (e.g., plug and chug). However, solving a novel or more difficult problem may be classified as constructive if a student has to rederive an equation, or reconceptualize certain components of a problem, such as reconceptualize two masses of a physics problem (one block on top of another block) as a single compound body. (p. 225)

Due to this nuance of constructive activities in the classroom, items were designed to measure these unobservable mental constructions that students may do in some classroom activities (for an example, see items 6c and 7a of section 3 in Appendix A). Items were also developed for observable constructive activities (see items 13 and 14 of section 2 in Appendix A).

A second main difficulty in writing items for the ICAP framework was designing items that could be used across all STEM disciplines. The item writers hoped that if the SCAEI could measure across science, engineering, and mathematics classrooms, it could begin to describe differences between STEM classrooms in terms of how students self-perceive the frequency of classroom activities. However, given the diverse disciplinary focuses across the STEM disciplines, it was a challenge to ensure that all discipline specific habits of mind were addressed by the SCAEI. To address this challenge, the item writers had collective disciplinary expertise in science, engineering, and mathematics.

## Methods to Initially Assess the SCAEI

It is important to note that instrument development requires a rigorous and iterative process that involves qualitative and quantitative methods of quality assessment [12]. This rigorous and iterative process mitigates the possibility of erroneous theoretical conclusions from poor instrument development and/or use [12, 16, 17, 18]. Two research-confirmed methods were used to assess the SCAEI: an item alignment study and cognitive interviews. These methods for initially assessing the SCAEI are detailed below.

*Item Alignment Study.* In an alignment study, content experts assess how well an item's content matches either (1) what is intended to be measured (i.e., frequency of certain classroom activities) or (2) the intended dimension of the theoretical framework used for operationalization (i.e., making an item measures a unique dimension of the ICAP framework) [12]. In this study, content experts assessed item alignment using an ordinal response scale (see Table 2), as recommended [12]. For example, in the SCAEI, content experts assessed how well the item "how often do you summarize the professor's lecture in your own words in class" represents, and thus measures, "constructive" classroom activities. The content experts' assessments of each item were then used (in conjunction with cognitive interview data) to determine whether to keep the item as is, revise the item, or omit it from the item bank.

**Table 2**
*Content Expert Ordinal Rankings for Alignment of the SCAEI*

| |
|---|
| a. Strong Alignment - The item’s content ALIGNS STRONGLY with the indicated content |

domain (passive, active, constructive, or interactive).

b. Acceptable Alignment - The item's content ALIGNS with the indicated content domain (passive, active, constructive, or interactive).
c. Insufficient Alignment - The item's content ALIGNS SLIGHTLY with the indicated content domain (passive, active, constructive, or interactive), but other domains are more appropriate.
d. No Alignment - The item's content DOES NOT ALIGN with the indicated content domain (passive, active, constructive, or interactive).
e. Unable to Judge – The item's content is TOO DIFFICULT TO JUDGE its alignment to the indicated content domain (passive, active, constructive, or interactive)

This process was pivotal to the item development process for the SCAEI because some classroom activities are "difficult to classify and fall on the boundary between two [ICAP] categories" [2, p. 225]. While the item developers tried to be cognizant of this aspect of the ICAP framework, the item alignment study process with content experts offered outside evaluation of this effort. Further, verifying that content experts across the STEM disciplines perceived each item as measuring a common, unique dimension of the ICAP framework ensured the relevance and utility of the instrument [12, 19]. As recommended by DeVellis [12], content experts also (1) assessed item clarity/readability with respect to the intended population (i.e., STEM undergraduate students) and (2) given the opportunity to recommend additional item content to tap important content of the ICAP framework that was not included in the SCAEI. These methods of content expert review are more rigorous than what is often employed in instrument development since expert reviews (i.e., item alignment studies) often make no effort to assess consistency or apply consistent criteria across different questionnaires [20].

To administer the alignment study, a digital version of the SCAEI was stored in Qualtrics and disseminated in-person to $N = 5$ content experts specializing in the following disciplines: one in engineering education, one in environmental engineering, one in chemistry, and two in mathematics. An interviewer was present as the content experts assessed the alignment of each item to a unique dimension of the ICAP framework; the interviewer asked follow-up questions and took field notes during the interviews. For example, if a content expert indicated less than "strong alignment" on any item, follow-up questions determined which other dimensions (if any) of the ICAP framework the content expert thought were also appropriate for the item and why.

*Cognitive Interviews.* Cognitive interviews [12, 21, 22, 23] are also recommended for initially assessing instruments being developed. Cognitive interviews consist of the target population responding to items developed for an instrument and explaining what they perceived the item to be about and how they formulated their response to the item (i.e., think-alouds). Cognitive interviews are commonly conducted to detect the presence of any construct-irrelevant variance that might be associated with items [21, 22, 23]. Construct-irrelevant variance is variance introduced from extraneous sources that is not related to what is being measured, which degrades measurement accuracy and precision [24]. The source of construct-irrelevant variance is sometimes due to lack of interpretability or readability of the items within the intended population [25]. Interpretability refers to making sure the items are interpreted as intended so that responses given to the items measure what was intended. Readability refers to making sure the items words are appropriate for the population (i.e., making sure that members of the population know what each word means). As an example, since the SCAEI is meant to function across all STEM undergraduates, using specialized or technical language specific to just one discipline in the items may introduce readability problems.

Methods for cognitive interviews included think-alouds [21, 23] and follow-up verbal probing [21, 23]. To administer the cognitive interviews, the items were loaded into a PowerPoint presentation which was used to display the items to each participant. Two interviewers co-conducted all but one of the $N$ = 5 cognitive interviews with individual engineering students. Both interviewers had a background in education research and one had expertise in instrument development. Interviewees were asked to read each item and voice their thoughts as they formulated their response to the item. Follow-up verbal probing was then conducted by the interviewers when construct-irrelevant variance might be present to ascertain its source if possible. If an item seemed problematic during the think-aloud, it was flagged for closer inspection and field notes were taken as to the nature of the problem. The interviewers later categorized items as problematic either due to (1) interpretability or (2) readability issues (i.e., non-related construct interference) using field notes. The cognitive interview protocol is included in Appendix B.

*Mixed Methods Design.* A concurrent mixed methods study [26, 27] was implemented to simultaneously gather quantitative data from content experts across STEM education and qualitative data from engineering students. Triangulation [27, 28] was used to identify problematic items by using both qualitative and quantitative data, which leads to a more informed conclusion from the data [27, 28]. For this study, both sets of data (i.e., the alignment study data and the cognitive interview data) were used to simultaneously evaluate the items for omission or revision from the SCAEI. If an item was problematic across multiple data sources (i.e., across content experts and multiple students), it was omitted. Items that seemed problematic in only one data source were considered individually for omission or revision.

Results

Results for the item alignment study are presented first followed by results for the cognitive interviews. Where applicable, results from the cognitive interviews are used to inform results from the item alignment study (and vice-versa), an affordance offered by the concurrent mixed-methods design [26, 27].

*Results for item alignment study.* Results indicated that one item should be removed for lack of alignment to the ICAP framework. Four other items were "flagged" but were not seen as problematic due to results from the cognitive interviews. Details follow for each item category ("passive" items, "active" items, etc.).

Across the "passive" and "active" items (see Appendix A), all content experts unanimously indicated either "acceptable alignment" or better. No further data for these items was evident in the field notes.

Three "constructive" items were flagged as having less than unanimous "acceptable alignment" (i.e., at least one content expert chose less than "acceptable alignment"). The first flagged "constructive" item (item 4c from Section 3) pertains to students making their own notes along with preexisting notes in class. The one content expert who flagged this item commented that some faculty make available their PowerPoint slides (i.e., preexisting notes) with blank spots (i.e., handouts with blanks) and that, in this context, this activity should not be categorized as "constructive" according to the ICAP framework. The second flagged "constructive" item (item 5b from Section 3) pertains to connecting course material to physical objects used in class. The concern was with student interpretation; that is, when we ask students to connect observations from a physical object to general course content knowledge, can we be sure that it is done at cognitive engagement level that is truly constructive? Concerns for both of these items center on student interpretation rather than the actual item alignment to the ICAP framework. As such, cognitive interview data for these items were considered to determine if such student interpretation problems existed (they did not). Hence, these items were seen as functioning as intended. The third flagged "constructive" item (item 8b from Section 3) pertained to writing short-answers to questions by applying old information to new ideas. One content expert flagged this item and argued that the wording of the question does not preclude the possibility that the partner referenced in the item did not contribute to writing the short-answer. Due to this, the item could be either "constructive" or "interactive". Edits to the item to preclude the "interactive" possibility were not seen as feasible, so the item was omitted.

Two "interactive" items (1a and 2a) were flagged as having "insufficient alignment." In both cases, two content experts expressed concerns that students may consider just getting feedback on "if they did the problem right" as pertaining to these questions. Again, these are issues related to student interpretations of the items, so the cognitive interview data was assessed for any similar discrepancies; however, none seemed apparent. Hence, these items were seen as functioning as intended.

As for overall functioning and validity of the SCAEI, all of the content experts stated that the SCAEI would be informative for guiding self-reflection on their own teaching. Some of the content experts also said they could see themselves using the SCAEI for education research purposes.

During the item alignment study, one content expert initially interpreted the "active" dimension of the ICAP framework as "active learning." When this mismatch was identified, the interviewer redirected the content expert to the definition for "active" as it relates to the ICAP framework. Further mismatch between the content expert's review of the SCAEI and ICAP framework was not detected.

*Results for cognitive interviews.* There were some items that produced unexpected results in the cognitive interviews even though they had functioned well in the alignment study (i.e., unanimous "acceptable alignment" or higher). These results are of a varied nature and are presented individually below.

One student considered a computer as a "physical object" that was used in class. This could prove problematic for items referring to physical objects (e.g., item 5 in section 3) because this may not align with what users of the instrument (i.e., STEM education researchers or practitioners) may consider in this question. More cognitive interviews will be conducted to see if other students have similar interpretations of the "physical object" phrase.

Another student found difficulty in the phrase "work with physical objects in class" (see item 5, section 1). He commented that he never "worked" with the physical objects in class, but he did "observe or hold" them. For this item, it was intended for “viewing” items to be interpreted as "working with" physical objects in class. Further cognitive testing will verbally probe whether other students have similar interpretations to this item.

Also, some “passive” items asked students if they just sit and wait for the instructor to provide answers. Some students indicated that they would at times "wait for the answer from the instructor or a classmate" after they had tried solving the problem on their own (i.e., doing more than just listening and waiting for an answer to the problem, therefore not passive). These items need to be edited to account for and exclude such a situation if they are to be categorized as "passive." This may be important for engineering educators to note. Students who hit a "roadblock" did wait for the instructor to provide the answer, but they were otherwise engaged with the content prior to "just sitting."

One student mentioned that instead of "in class," the phrase "during class" would help them to remember to only consider events that happened inside of a class period instead of also considering events related to the items but that occurred outside of class (e.g., during study groups). All instances of "in class" will be edited to "during class" and another round of cognitive interviews will assess if this change impacts students' interpretations of the items.

One unexpected result was that all of the engineering students were not comfortable with the wording of item 2 in section 2 ("defend or argue" with the instructor). Many indicated that this phrase had a "negative connotation." Although this interpretation varies slightly from the item developers’ intention, it is common for instrument developers to leave items such as this intact [3]. There are two reasons for not editing the item. First, the students’ interpretation does not alter the items ICAP alignment (i.e., both interpretations fit the same category). Second, items such as these help separate respondents (i.e., create variation between respondents' measures). This is advantageous to instrument development and is akin to replacing measures produced with a ruler marked in centimeters with measures from one marked in millimeters.

Discussion and Implications

The data from the alignment study and the cognitive interviews are important steps in the development of SCAEI. The data presented here form an initial step of the iterative revision and vetting process. The input provided by the engineering students in the cognitive interviews highlights how social structures may impede engagement in behaviors the instructor may wish to promote in the classroom (i.e., defending ideas and scientific/mathematical argumentation). Being cognizant of how these students interpreted and responded to such survey items is important for both instrument design and item testing and also illustrates how important these processes are to instrument development.

With respect to the theoretical framework, the fact that one content expert initially interpreted the "active" dimension of ICAP as "active learning" has implications for future uses the SCAEI. Care must be given by education researchers and practitioners to understand exactly what the

SCAEI measures and how those measures should be used. Others may have similar misunderstandings about the SCAEI, so care must be given to make the distinction between "active learning" and "active" with respect to ICAP. Chi and Wylie [2] were very cognizant of this in their presentation of the ICAP framework, and this cognizance must persist for the SCAEI to ensure the validity of the instrument (i.e., proper use and interpretation of measures produced by the SCAEI). However, this study has found other initial sources of validity for the SCAEI, including: (1) seemingly accurate interpretations of the items by engineering students, (2) accurate alignment of what the instrument is measuring as evaluated by content experts, and (3) support of the instrument and planned intended use of the instrument by education researchers and practitioners. The initial steps for validating the SCAEI presented here, steps which are often overlooked or ignored by instrument developers [12], have provided valuable information for the development of the SCAEI.

These results also indicated social and behavioral context that engineering instructors should consider when planning classroom activities. Specifically, the engineering students perceived “arguing” or “defending” ideas as something that is disrespectful to the instructor. If an instructor is trying to engage students in this type of professional discourse, it may be important to have a discussion about what it means to scientifically defend a position. Furthermore, students indicated that they would become passive if they hit a “roadblock” while problem-solving. As a critical next step for these students, instructors should consider the incorporation of methods that facilitate students developing the capability to overcome these roadblocks.

## Conclusion and Next Steps

The SCAEI is being constructed to measure students’ perceptions of their experiences in the classroom in relation to the ICAP framework. Such measures are needed [10], and the consistency in item alignment for the majority of items across content experts provides content, construct, and face validity evidence for the SCAEI. The possibilities for the uses of the instrument include: (1) providing faculty with tools to understand how students interact in and perceive STEM courses, (2) providing evaluation data for faculty concerning their teaching practices, (3) correlating student perceptions of the classroom environment to student learning outcomes, and (4) identifying strengths of different instructors and leveraging that strength to improve instructional strategies either within or across other departments.

After revisions of the SCAEI are complete, another round of the alignment study and cognitive interviews should affirm that the revisions have improved item alignment to the ICAP framework and/or accurate interpretation of the items by STEM students. Analysis of how the instrument quantitatively functions within the intended population should then be assessed.

## Acknowledgements

The authors are grateful for support provided by the National Science Foundation grant DUE 1347817. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References


[1] Chi, M. T. H., (2009). Active-Constructive-Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*, 73-105.

[2] Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes, *Educational Psychologist, 49*(4), 219-243.

[3] Kramarski, B., & Dudai, V. (2009). Group-metacognitive support for online inquiry in mathematics with differential self-questioning. *Journal of Educational Computing Research*, *40*, 377–404.

[4] Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74.

[5] Hausmann, R. G. M., van de Sande, B., & VanLehn, K. (2008). Shall we explain? Augmenting learning from intelligent tutoring systems and peer collaboration. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent tutoring systems* (pp. 636–645). Amsterdam, The Netherlands: IOS.

[6] Menekse, M., Stump, G., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education, 102*, 346–374.

[7] Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415.

[8] Hora, M. T., & Ferrare, J. J. (2013). Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching. Journal of the Learning Sciences, 22(2), 212-257.

[9] Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. CBE-Life Sciences Education, 12(4), 618-627.

[10] American Association for the Advancement of Science. (2012). Describing & measuring undergraduate STEM teaching practices: A report from a national meeting on the measurement of undergraduate science, technology, engineering, and mathematics (STEM) teaching. Available at http://ccliconference.org/files/2013/11/Measuring-STEM-Teaching-Practices.pdf

[11] Prince, M. (2004). Does active learning work? A review of the research. Journal of engineering education, 93(3), 223-231.

[12] DeVellis, R. F. (2012). *Scale development: Theory and applications.* Thousand Oaks, California: SAGE Publications, Inc.

[13] de Ayala, R. J. (2009). *The theory and practice of item response theory.* New York, New York: Guildford Press.

[14] Smith, T. W., (2004). Developing and evaluating cross-national survey instruments. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 431-452). Hoboken, New Jersey: John Wiley & Sons, Inc.

[15] Alwin, D. F., (2010). How good is survey measurement? Assessing the reliability and validity of survey measures. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 405-434). Wagon Lane, Bingley, United Kingdom: Emerald Group Publishing.

[16] Benson, J., & Clark, F. (1982). A guide for instrument development and validation. *The American Journal of Occupational Therapy, 36,* 789-800.

[17] Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-26). Hillsdale, New Jersey: Lawrence Erlbaum.

[18] Linn, R. L. (2006). *The standards for educational and psychological testing: Guidance in test development.* In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-26). Hillsdale, New Jersey: Lawrence Erlbaum

[19] Messick, S. (1990). *Validity of test interpretation and use* (ETS Research Report 90-11). Retrieved from http://files.eric.ed.gov/fulltext/ED395031.pdf.

[20] Tourangeau, R. (2004). Experimental design considerations for testing and evaluating questionaires. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 209-224). Hoboken, New Jersey: John Wiley & Sons, Inc.

[21] Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 263-313). Wagon Lane, Bingley, United Kingdom: Emerald Group Publishing.

[22] Conrad, F., & Blair, J. (1996). *From Impressions to data: Increasing the objectivity of cognitive interviews*. Paper presented at the meeting of the ASA Section on Survey Research Methods, Alexandria, Virginia. Retrieved from http://www.bls.gov/ore/pdf/st960080.pdf

[23] Willis, G. B. (1999). *Cognitive interviewing: A "How To" guide*. Paper presented at the 1999 Meeting of the American Statistical Association, Baltimore, Maryland.

[24] McMillan, J. H., & Schumacher, S. (2010). *Research in education: Evidence-based inquiry*. Upper Saddle River, New Jersey: Pearson Education, Inc.

[25] DeMaio, T. J., & Landreth, A. (2004). Do different cognitive interview techniques produce different results? In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 361-384). Hoboken, New Jersey: John Wiley & Sons, Inc.

[26] Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* Upper Saddle River, New Jersey: Merrill Prentice Hall.

[27] Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 209-240). Thousand Oaks, California: Sage Publications, Inc.

[28] Erzberger, C., & Kelle, U. (2003). Making inferences in mixed methods: The rules of integration. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 457-488). Thousand Oaks, California: Sage Publications, Inc.

**Appendix A**
**Student Class Activity and Engagement Instrument**

**General Instructions**
Thank you for participating in this study! We are hoping to learn about what you do in class so that we can better understand the classroom environment. Your answers here will be about the frequency in which you do some various classroom activities. There are no right or wrong answers.

**Section 1 - For the student version of the instrument, all items in this section have "yes" or "no" response options. Student responses in this section determine which of the corresponding questions from Section 3 are displayed to students to help alleviate response burden. When piloting with students, question order will be randomized. Content experts did not evaluate these items for alignment to the ICAP framework since these items do not clearly align with any specific dimension without further clarification (which is in essence what the questions in Section 3 further specify).**

**Instructions:** Below are a series of questions about various things you may or may not do in class. Please answer yes or no for each question as best you can. There are no right or wrong answers.

1. Do you problem solve (i.e., think critically) during class?
2. Do you ask the instructor questions in class?
3. Do you ever work on course content in class with classmates?
4. Do you ever use preexisting notes in class (i.e., professor handouts, notes from reading, etc.)?
5. Do you work with physical objects (i.e., rock samples, biological specimens, etc.) in class?
6. Do you ever use clickers in class?
7. Do you ever write/construct short-answers to questions by yourself in class?
8. Do you ever write/construct short-answers to questions with a partner in class?
9. Do you ever respond to questions asked by the instructor in class?
10. Do you ever respond to questions asked by a classmate (questions directed either to you or the instructor) in class?
11. Do you take notes in class?

**Section 2 - For the student version of the instrument, all items in this section employ a 4-point Likert frequency scale: "never", "rarely", "sometimes", and "frequently". Text below in parenthesis (i.e., the corresponding dimension of the ICAP framework for each item) is not displayed to students. Questions p and q have follow-up open-response questions to help identify if further item development needs to occur. When piloting with students, question order will be randomized. For the content expert version of the instrument, all items in this section employ a 5-point Likert scale (see Table 2).**

**Instructions:** Below are a series of questions about various things you might do in class. For each question, please indicate the level of frequency which best matches how often you do this in <COURSE> as best you can. There are no right or wrong answers.

How often do you:

1. (Interactive) Get input on your ideas from the instructor in class?
2. (Interactive) Defend or argue your idea to the instructor in class?
3. (Interactive) Get input on your ideas from a classmate(s) in class?
4. (Interactive) Defend or argue your idea to a classmate(s) in class?
5. (Active) Take notes verbatim in class?
6. (Active) Copy solution steps verbatim in class?
7. (Active) Highlight your notes in class?
8. (Active) Repeat or rehearse things needed for tests or quizzes in class?
9. (Constructive) Take notes in your own words in class?
10. (Constructive) Reflect out-loud to yourself in class?
11. (Constructive) Summarize the professor's notes in your own words in class?
12. (Constructive) Summarize the professor's lecture in your own words in class?
13. (Constructive) Make your own diagrams to help you understand material in class?
14. (Constructive) Draw concept maps in class?
15. (Constructive) Compare and contrast new ideas to old knowledge in class?
16. (Passive) Just listen without doing anything else (even note taking) in class?
17. (Passive) Watch a multimedia presentation (i.e., video, PowerPoint presentation, etc.) without doing anything else (even note taking) in class?

TO BE EDITED - Change "in class" to "during class".

**Section 3 - For the student version of the instrument, all items in this section employ a 4-point Likert frequency scale: "never", "rarely", "sometimes", and "frequently". Text below in parenthesis (i.e., the corresponding dimension of the ICAP framework for each item) is not displayed to students. Questions p and q have follow-up open-response questions to help identify if further item development needs to occur. When piloting with students, question order will be randomized. For the content expert version of the instrument, all items in this section employ a 5-point Likert scale (see Table 2).**

1. You said that you sometimes problem solve (i.e., think critically) in class. When you are given time to problem solve in class:
    a. (Interactive) How often do you work with others to better understand course material?
    b. (Constructive) How often do you learn something new to better understand the course material?
    c. (Active – rehearsal) How often do you use a process that you are very familiar with to solve these problems?
    d. (Passive) How often do you just wait for the answer from the instructor or a classmate? - TO BE EDITED
2. You said that you sometimes ask the instructor questions in class. When you ask the instructor questions:
    a. (Interactive) How often is it to further your understanding of the material?
    b. (Non-ICAP) How often is it to just clarify something you misheard or to get the instructor to repeat what they said?
    c. [OPEN RESPONSE QUESTION] Do you ask questions for reasons not listed above? If so, for what reasons do you ask questions in class?
3. You said that you sometimes work on course content in class with classmates. When you work on course content in class with classmates:
    a. (Interactive) How often do classmates provide input on your ideas in class?
    b. (Interactive) How often do you defend or argue your idea to a classmate in class?
    c. (Interactive) How often do you ask questions to your classmate to further your understanding of the material?
    d. (Active) How often do you repeat information to each other?
    e. (Active) How often is it just practicing techniques needed for homework, tests, or quizzes?
    f. (Passive) How often do you just wait for the professor to move on without thinking critically with your classmate(s)?
    g. FOLLOW UP QUESTION: Do you ask questions for reasons not listed above? If so, what are they?
4. You said that you sometimes use preexisting notes (i.e., professor handouts, notes from reading, etc.) in class. When you use preexisting notes:
    a. (Active) How often do you highlight these notes in class?
    b. (Passive) How often do you follow along with these notes without writing on them? - TO BE EDITED
    c. (Constructive) How often do you make your own notes along with the preexisting notes?
5. You said that you sometimes work with physical objects (i.e., rock samples, biological specimens, etc.) in class. When you work with physical objects:
    a. (Active) How often do you just look at the object and not connect it to the course material?
    b. (Constructive) How often do you look at the object and connect it to the course material?
6. You said that you sometimes use clickers in class. When you use clickers in class:
    a. (Non-ICAP) Is it to take attendance only?
    b. (Active) How often do you just have to recall information to answer the clicker question?
    c. (Constructive) How often do you have to synthesize multiple sources of information to answer the question?
    d. (Interactive) How often do you work with other people to answer the clicker question?
7. You said that you sometimes write/construct short-answers to questions by yourself in class. When you write/construct short-answers to these questions by yourself in class:
    a. (Constructive) How often do these questions require you to apply old knowledge to new ideas to determine the correct answer?
    b. (Active – rehearsal) How often do just have to remember information to answer these questions?
    c. (Passive) How often do you just wait for the answer from the instructor or a classmate? - TO BE EDITED

8. You said that you sometimes write/construct short-answers to questions with a partner in class. When you write/construct short-answers to these questions with a partner in class:
    a. (Interactive) How often does your partner's help with the question provide ideas or insight that you do not think you could have come up with alone?
    b. (Constructive) How often do these questions require you to apply old knowledge to new ideas to determine the correct answer? - ITEM OMITTED
    c. (Active – rehearsal) How often do just have to remember information to answer these questions?
    d. (Passive) How often do you just wait for the answer from the instructor or a classmate? - TO BE EDITED
9. You said that you sometimes respond to questions asked by the instructor in class. When you respond to the instructor's questions in class:
    a. (Active – rehearsal) How often do just have to remember information to answer these questions?
10. You said that you sometimes respond to questions asked by a classmate (questions directed either to you or the instructor) in class. When you respond to questions asked by a classmate in class:
    a. (Interactive) How often does your classmate's question provide ideas or insight that you had not thought of before?
    b. (Constructive) How often do these questions require you to apply old knowledge to new ideas to determine the correct answer?
    c. (Active – rehearsal) How often do just have to remember information to answer these questions?

TO BE EDITED - Change "in class" to "during class".

**Appendix B**
**Student Cognitive Interview Protocol**

**Directions**
Read the warm-up exercise to the participant. Help acquaint the participant with the think-aloud process. Then, give the participant the sample question. Finally, use the interview protocol at the bottom to have the participant respond to each item of the instrument.

**Warm-up Exercise**
First, I am going to do a warm-up with you to acquaint you with what we will be doing for the interview. "Try to visualize the place where you live, and think about how many windows there are in that place. As you count up the windows, tell me what you are seeing and thinking about" (Willis, 1999).

**Sample Question**
Now I'm going to show you a sample question. <Display the sample item "how often did you correct an error you made after in-class feedback from your instructor?">

1. As you read this question, tell me what you are thinking about.
2. Can you repeat the question in your own words?
    a. {If misalignment appears to exist, ask this follow-up question.} What was it about the question that made you think of it that way?
    b. {Optional Probe} Is there anything unclear about the question? If so, what is unclear and why?
3. Do you think any of your college peers would interpret the question different from you?
    a. {If they answer yes, ask these follow-up questions.} How do you think they would interpret it different?
    b. Why do you think they would interpret it that way?

**Interview Protocol**
Now I am going to show you a list of questions pertaining to sustainability. We will go through the same process with all of these items. Your feedback will help me identify items that may be problematic, so please feel comfortable to provide critical feedback. Do you have any questions for me before we start?

<Begin sample items>

1. As you read this question, tell me what you are thinking about.
2. Can you repeat the question in your own words?
    a. {If misalignment appears to exist, ask this follow-up question.} What was it about the question that made you think of it that way?
    b. {Optional Probe} Is there anything unclear about the question? If so, what is unclear and why?
3. Do you think any of your college peers would interpret the question different from you?
    a. {If they answer yes, ask these follow-up questions.} How do you think they would interpret it different?
    b. Why do you think they would interpret it that way?