APPROVAL OF PROTOCOL

August 20, 2013

Paula Lemons

706-542-3340
plemons@uga.edu

Dear Paula Lemons:

On 8/20/2013, the IRB reviewed the following submission:

<table>
<thead>
<tr>
<th>Type of Review</th>
<th>Initial Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title of Study</td>
<td>A community of enhanced assessment facilitates reformed teaching</td>
</tr>
<tr>
<td>Investigator</td>
<td>Paula Lemons</td>
</tr>
<tr>
<td>IRB ID</td>
<td>STUDY00000257</td>
</tr>
<tr>
<td>Funding</td>
<td>Name: NATIONAL SCIENCE FOUNDATION; Grant Office ID: not yet assigned, Funding Source ID: DUE 1347733,</td>
</tr>
<tr>
<td>Grant ID</td>
<td>not yet assigned;</td>
</tr>
<tr>
<td>IND, IDE, or HDE</td>
<td>None</td>
</tr>
</tbody>
</table>
| Documents Reviewed | • Consent form_Faculty, Category: Consent Form;  
                      • Consent form_Administrators, Category: Consent Form; |

The IRB approved the protocol through the exempt (administrative) review procedure from 8/20/2013 until 8/19/2018.

To document consent, use the consent documents that were approved and stamped by the IRB. Go to the Documents tab to download them.

In conducting this study, you are required to follow the requirements listed in the Investigator Manual (HRP-103).

Sincerely,

Larry Nackerud, PhD  
University of Georgia  
Institutional Review Board Chairperson
1. **Title of study:**
   A community of enhanced assessment facilitates reformed teaching

2. **Short title:**
   Enhanced assessment facilitates reformed teaching

3. **Brief description:**
   For the past 20 years, NSF has funded millions of dollars in projects to develop curriculum that promotes reformed teaching in college science classes. They have also focused on disseminating these curricula. Unfortunately, recent research shows problems with the dissemination model. For example, there is a well-documented fall-off of faculty who have tried and later abandoned a wide variety of reformed teaching practices.

   This project aims to promote widespread adoption and sustained use of constructed-response assessments in college science courses through the formation of local Faculty Learning Communities (FLCs) at UGA and other institutions. These FLCs will focus on developing reflective teachers and shared vision among individual faculty and administrators. Focusing in these areas has been shown to lead to successful reform in educational practices.

   The research proposed in this protocol aims to understand the experiences of faculty who participate in FLCs, including the patterns and trends of their participation, the systemic structures in their departments, institutions, and disciplines that influence their use of constructed-response assessments, and their ideas and conceptions about teaching and learning.

4. **Principal investigator:**
   Paula Lemons

5. **Does the investigator have a financial interest related to this research?**
   - [ ] Yes  [ ] No

6. **Will an external IRB act as the IRB of record for this study?**
   - [ ] Yes  [ ] No

7. **Attach the protocol:** (include the investigator protocol and full sponsor protocol)
   
<table>
<thead>
<tr>
<th>Document</th>
<th>Category</th>
<th>Date Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>A community of enhanced assessment facilitates reformed teaching(1)</td>
<td>IRB</td>
<td>8/16/2013</td>
</tr>
</tbody>
</table>

   **Use one of these templates:**
View: SF: Funding Sources (not integrated with Grants)

**Funding Sources**

1. **Identify each organization supplying funding for the study:**

<table>
<thead>
<tr>
<th>Funding Source</th>
<th>Sponsor's ID</th>
<th>Grants Office ID</th>
<th>Attachments</th>
</tr>
</thead>
<tbody>
<tr>
<td>NATIONAL SCIENCE FOUNDATION</td>
<td>DUE 134773</td>
<td>not yet assigned</td>
<td>Collaborative Research: A community of enhanced assessment facilitates reformed teaching</td>
</tr>
</tbody>
</table>

View: SF: Study Team Members

**Study Team Members**

1. **Identify each additional person involved in the design, conduct, or reporting of the research:**

<table>
<thead>
<tr>
<th>Name</th>
<th>Roles</th>
<th>Financial Interest</th>
<th>Involved in Consent</th>
<th>E-mail</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taylor Adkins</td>
<td>CO-INVESTIGATOR/STUDENT</td>
<td>no</td>
<td>no</td>
<td><a href="mailto:tradkins@uga.edu">tradkins@uga.edu</a></td>
<td></td>
</tr>
<tr>
<td>Tessa Andrews</td>
<td>CO-INVESTIGATOR</td>
<td>no</td>
<td>no</td>
<td><a href="mailto:tandrews@uga.edu">tandrews@uga.edu</a> 706-542-3340</td>
<td></td>
</tr>
<tr>
<td>TRECHERIE Crumbs</td>
<td>SUB-INVESTIGATOR/STUDENT</td>
<td>no</td>
<td>no</td>
<td><a href="mailto:tcrombs@uga.edu">tcrombs@uga.edu</a></td>
<td></td>
</tr>
<tr>
<td>Ersta Ferryanto</td>
<td>SUB-INVESTIGATOR/STUDENT</td>
<td>no</td>
<td>no</td>
<td><a href="mailto:ersta@uga.edu">ersta@uga.edu</a></td>
<td></td>
</tr>
<tr>
<td>Alexandra Howell</td>
<td>SUB-INVESTIGATOR/STUDENT</td>
<td>no</td>
<td>no</td>
<td><a href="mailto:ahowell3@uga.edu">ahowell3@uga.edu</a></td>
<td></td>
</tr>
<tr>
<td>Jill Voreis</td>
<td>CO-INVESTIGATOR</td>
<td>no</td>
<td>yes</td>
<td><a href="mailto:jvoreis@uga.edu">jvoreis@uga.edu</a> 706-542-1334</td>
<td></td>
</tr>
</tbody>
</table>

View: SF: Study Scope

**Study Scope**

1. **Are there external sites where the investigator will conduct or oversee the research?**
   - Yes
   - No

2. **Does the study do any of the following:**
   - Specify the use of an approved drug or biologic?
   - Use an unapproved drug or biologic?
   - Use a food or dietary supplement to diagnose, cure, treat, or mitigate a disease or condition?
3. *Does the study do any of the following:*

- Evaluate the safety or effectiveness of a device?
- Use a humanitarian use device (HUD)?

(Yes) ☐ No

View: SF: External Sites

**External Sites**

1. *Identify each external site where the investigator will conduct or oversee the research:*

<table>
<thead>
<tr>
<th>Site</th>
<th>Contact</th>
<th>Phone</th>
<th>E-mail</th>
<th>External IRB Review</th>
<th>Rely on This IRB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michigan State University</td>
<td>Mark Urban-Lurain</td>
<td>(517) 432-2108</td>
<td><a href="mailto:urban@msu.edu">urban@msu.edu</a></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Ohio State University</td>
<td>Meghan Federer</td>
<td>614-648-3462</td>
<td><a href="mailto:federer.21@osu.edu">federer.21@osu.edu</a></td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>State University of New York</td>
<td>Ross Nehm</td>
<td>631-632-7247</td>
<td><a href="mailto:ross.nehm@stonybrook.edu">ross.nehm@stonybrook.edu</a></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>University of Colorado Boulder</td>
<td>Jennifer Knight</td>
<td>303-735-1949</td>
<td><a href="mailto:knight@colorado.edu">knight@colorado.edu</a></td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>University of Maine</td>
<td>Michelle Smith</td>
<td>207-581-2604</td>
<td><a href="mailto:michelle.k.smith@maine.edu">michelle.k.smith@maine.edu</a></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>University of South Florida</td>
<td>Luanna Prevost</td>
<td>706-207-8221</td>
<td><a href="mailto:prevost@usf.edu">prevost@usf.edu</a></td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Western Michigan University</td>
<td>Mary Anne Sydlik</td>
<td>269-387-5393</td>
<td><a href="mailto:maryanne.sydlik@wmich.edu">maryanne.sydlik@wmich.edu</a></td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>

View: SF: Recruitment Materials

**Consent Forms and Recruitment Materials**

1. **Consent forms:** (include an HHS-approved sample consent document, if applicable)

<table>
<thead>
<tr>
<th>Document</th>
<th>Category</th>
<th>Date Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Consent form_Administrators(2)</td>
<td>Consent Form</td>
<td>8/19/2013</td>
</tr>
<tr>
<td>View Consent form_Faculty_Modified (4)</td>
<td>Consent Form</td>
<td>11/27/2013</td>
</tr>
<tr>
<td>View Consent form_Faculty(2)</td>
<td>Consent Form</td>
<td>8/19/2013</td>
</tr>
</tbody>
</table>

Refer to the following templates and instructional documents:

- Consent Template - Consent Form (with signature)
- Consent Template - Consent Cover Letter (no signature)
- Policy and Procedure: Informed Consent Process for Research
- Policy and Procedure: Documentation of Informed Consent

2. **Recruitment materials:** (add all material to be seen or heard by subjects, including ads)

<table>
<thead>
<tr>
<th>Document</th>
<th>Category</th>
<th>Date Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Recruitment email_faculty(1)</td>
<td>Recruitment Materials</td>
<td>8/16/2013</td>
</tr>
<tr>
<td>View Recruitment email_administrators(1)</td>
<td>Recruitment Materials</td>
<td>8/16/2013</td>
</tr>
</tbody>
</table>
View: IRB_Project_Funding_Details

Project Funding Details

1. **Identify below if the study is/will be supported in whole or in part by external or internal funds.**
   
   **Funding Status:**
   
   Pending

2. **Name of Project and/or Project PI if different from this IRB Application.**
   
   Collaborative Research: A community of enhanced assessment facilitates reformed teaching

3. **Describe the scope of this IRB Application compared to the grant proposal (sub-award or statement of work) related to this project.**
   
   **Scope of Application:**
   
   All human subjects activities in the grant proposal are covered in this IRB Application.

4. **If required above to provide additional information, please do so in the space below.**

View: NonUGA_IRB_Collaborators

Non-UGA Collaborators

1. **Identify non-UGA personnel who will be engaged in the conduct of human research.**

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Alexandria Mazur</td>
<td><a href="mailto:mazorale@msu.edu">mazorale@msu.edu</a></td>
<td>Michigan State University</td>
</tr>
<tr>
<td>View Andrea Espina</td>
<td><a href="mailto:prevost@usf.edu">prevost@usf.edu</a></td>
<td>University of South Florida</td>
</tr>
<tr>
<td>View Dania Malik</td>
<td><a href="mailto:dania.malik@stonybrook.edu">dania.malik@stonybrook.edu</a></td>
<td>SUNY Stonybrook</td>
</tr>
<tr>
<td>View Erin Vinson</td>
<td><a href="mailto:erin.vinson@maine.edu">erin.vinson@maine.edu</a></td>
<td>University of Maine</td>
</tr>
<tr>
<td>View Jennifer Knight</td>
<td><a href="mailto:knight@colorado.edu">knight@colorado.edu</a></td>
<td>University of Colorado at Boulder</td>
</tr>
<tr>
<td>View Jeremy Rentsch</td>
<td><a href="mailto:Jere1891@colorado.edu">Jere1891@colorado.edu</a></td>
<td>University of Colorado at Boulder</td>
</tr>
<tr>
<td>View John Merrill</td>
<td><a href="mailto:merrill3@msu.edu">merrill3@msu.edu</a></td>
<td>Michigan State University</td>
</tr>
<tr>
<td>View Karen Pelletreau</td>
<td><a href="mailto:karen.pelletreau@maine.edu">karen.pelletreau@maine.edu</a></td>
<td>University of Maine</td>
</tr>
<tr>
<td>View Kathleen Dinota</td>
<td><a href="mailto:kdinota@eischools.org">kdinota@eischools.org</a></td>
<td>SUNY - University at Stony Brook</td>
</tr>
<tr>
<td>View Kelli Hayes</td>
<td><a href="mailto:prevost@usf.edu">prevost@usf.edu</a></td>
<td>University of South Florida</td>
</tr>
<tr>
<td>View Kirsti Martinez</td>
<td><a href="mailto:prevost@usf.edu">prevost@usf.edu</a></td>
<td>University of South Florida</td>
</tr>
<tr>
<td>View Luanna Prevost</td>
<td><a href="mailto:prevost@usf.edu">prevost@usf.edu</a></td>
<td>University of South Florida</td>
</tr>
<tr>
<td>View Margaurete Romero</td>
<td><a href="mailto:mromero@mail.usf.edu">mromero@mail.usf.edu</a></td>
<td>University of South Florida</td>
</tr>
<tr>
<td>View Mark Urban-Lurain</td>
<td><a href="mailto:urban@msu.edu">urban@msu.edu</a></td>
<td>Michigan State University</td>
</tr>
<tr>
<td>View Mary Anne Sydlik</td>
<td><a href="mailto:mary.sydlik@wmich.edu">mary.sydlik@wmich.edu</a></td>
<td>Western Michigan University</td>
</tr>
<tr>
<td>View Meghan Federer</td>
<td><a href="mailto:federer.21@buckeyemail.osu.edu">federer.21@buckeyemail.osu.edu</a></td>
<td>Ohio State University</td>
</tr>
<tr>
<td>View Michelle Smith</td>
<td><a href="mailto:michelle.k.smith@maine.edu">michelle.k.smith@maine.edu</a></td>
<td>University of Maine System</td>
</tr>
<tr>
<td>View Miha Park</td>
<td><a href="mailto:parkmh1@msu.edu">parkmh1@msu.edu</a></td>
<td>Michigan State University</td>
</tr>
<tr>
<td>View Rosa Moscarella</td>
<td><a href="mailto:moscarel@msu.edu">moscarel@msu.edu</a></td>
<td>Michigan State University</td>
</tr>
<tr>
<td>View Ross Nehm</td>
<td><a href="mailto:ross.nehm@stonybrook.edu">ross.nehm@stonybrook.edu</a></td>
<td>SUNY - University at Stony Brook</td>
</tr>
<tr>
<td>View Scott Merrill</td>
<td><a href="mailto:scott.merrill@umit.maine.edu">scott.merrill@umit.maine.edu</a></td>
<td>University of Maine</td>
</tr>
</tbody>
</table>
*Attached documents may be reviewed under the "Documents" tab of the study home page.

*Submit an Individual Investigator Agreement for all study personnel affiliated with an institution that does not have an assurance with the Office for Human Research Protections or OHRP (typically, local schools, private doctors’ clinics). Upload agreement(s) under "Supporting Documents."

View: Student Research

**Student Research**

1. **Complete this section for each student co-investigator listed on the Study Team Members page who will use this research project for thesis/dissertation.** Important Note: The IRB recommends submission for IRB review only after the appropriate committee has conducted the necessary scientific review and approved the research proposal.

   **Student Name**  **Student Research Purpose**  **Student Committee Approval**  **(IRB Student Research)**  **Other**

   There are no items to display

View: IRB CITI Training List

**Study Team Members CITI Training Records**

**Principal Investigator (PI):** Paula Lemons

**Job Title:** ASSOCIATE PROFESSOR

**PI's Completed Training Courses:**

<table>
<thead>
<tr>
<th>Training Course</th>
<th>Expiration Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Research: Social &amp; Behavioral Research</td>
<td>7/31/2019</td>
</tr>
</tbody>
</table>

**Study Team Members:**

<table>
<thead>
<tr>
<th>Name</th>
<th>Completed Training Courses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taylor Adkins</td>
<td>Training Course</td>
</tr>
<tr>
<td></td>
<td>Human Research: Social &amp; Behavioral Research</td>
</tr>
<tr>
<td></td>
<td>3/18/2018</td>
</tr>
<tr>
<td>Tessa Andrews</td>
<td>Training Course</td>
</tr>
<tr>
<td></td>
<td>Human Research: Social &amp; Behavioral Research</td>
</tr>
<tr>
<td></td>
<td>6/4/2017</td>
</tr>
<tr>
<td></td>
<td>Human Research: Social &amp; Behavioral Research - Children</td>
</tr>
<tr>
<td></td>
<td>6/5/2017</td>
</tr>
<tr>
<td></td>
<td>Human Research: Social &amp; Behavioral Research - Internet</td>
</tr>
<tr>
<td></td>
<td>6/5/2017</td>
</tr>
<tr>
<td>TRECHERIE Crumbs</td>
<td>Training Course</td>
</tr>
<tr>
<td></td>
<td>Human Research: Social &amp; Behavioral Research</td>
</tr>
<tr>
<td></td>
<td>10/1/2019</td>
</tr>
<tr>
<td></td>
<td>Human Research: Social &amp; Behavioral Research - Internet</td>
</tr>
<tr>
<td></td>
<td>10/1/2019</td>
</tr>
<tr>
<td>Ersta Ferryanto</td>
<td>Training Course</td>
</tr>
<tr>
<td></td>
<td>Human Research: Social &amp; Behavioral Research</td>
</tr>
<tr>
<td></td>
<td>1/7/2019</td>
</tr>
<tr>
<td></td>
<td>Human Research: Social &amp; Behavioral Research</td>
</tr>
<tr>
<td></td>
<td>1/7/2019</td>
</tr>
</tbody>
</table>
1. **Provide a general description of the targeted participants** *(e.g., healthy adults from the general population, children enrolled in an after-school program, adolescent females with scoliosis)*, and **indicate the targeted gender** *(males, females, or both)*, **age/age range** *(e.g., 18 years old and above)*, and **total maximum number (or range) of participants needed to complete the study**. "Total Number/Range" should include the anticipated number of those who will give consent but fail screening (if research involves pre-screening participants) and other reasons for attrition *(e.g., withdrawals, or termination by researchers)*.

<table>
<thead>
<tr>
<th>Targeted Population</th>
<th>Targeted Gender</th>
<th>Age or Age Range</th>
<th>Total Number / Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>View College faculty and administrators</td>
<td>No gender is targeted</td>
<td>Any age over 18</td>
<td>20-30 faculty; 5-15 administrators</td>
</tr>
</tbody>
</table>

2. **Identify the inclusion and exclusion criteria.** If two or more targeted populations, identify eligibility criteria for each.

**Inclusion Criteria:**
Faculty participants in the Automated Analysis of Constructed Response project. Specifically members of Faculty Learning Communities associated with this project. Also, select administrators who oversee these faculty will be included. The administrators are those who have expressed their commitment to the project by providing letters of commitment to NSF for this project.

**Exclusion Criteria:**
No members of the project, as specified above, will be excluded.

3. **Describe how the eligibility based on the above inclusion/exclusion criteria will be determined** *(e.g., screening by phone, self-report via a screening questionnaire, self-select by reviewing eligibility criteria on recruitment materials, hospital records, school records, additional tests/exams)*.

**Eligibility Determination:**
The PI will have a list of all participants. The responses to the recruiting email will be cross-checked against the list.

4. **If the research will exclude a particular gender or minority group, please provide justification.**
Exclusion Justification:
N/A

5. Will participants receive any incentives for their participation (e.g., payments, gifts, compensation, reimbursement, services without charge, extra class credit)? If yes, please describe. If offering extra class credit, describe a comparable non-research alternative for receiving incentive.

Incentive Description:
Yes, participants will receive an incentive for their participation. Any members of the Faculty Learning Communities who participate in research will receive $1500 for each year they participate in research. Participants in the FLC who do not participate in research will not be compensated.

View: IRB_Sp_pops

Vulnerable and/or Special Populations

1. Click any/all that apply and attach applicable checklists to the Supporting Documents page.
   Population:
   UGA Psychology Research Pool/Other UGA students/employees
   
   UGA IRB Subpart B Checklist (Pregnant women, neonates, or fetuses)
   UGA IRB Subpart C Checklist (Prisoners)

2. Provide justification for including the group(s) checked above in this particular study.

   UGA employees are the target population for this study because they are the ones participating in the faculty professional development that this research investigates. Findings from UGA will be compared with findings from other institutions. Findings will inform future faculty professional development for UGA employees.

3. If there is a working relationship between any researchers and the participants (e.g., PI's own students or employees), describe below.

   N/A

4. Describe the safeguards to protect the rights and welfare of these participants and to minimize any possible coercion or undue influence. For example, amount of payment will be non-coercive for the financially disadvantaged, extra careful evaluations of participants’ understanding of the study, advocates to be involved in the consent process, or use flyers to recruit participants instead of directly approaching own staff or students.

   N/A

View: IRB Recruitment Procedures
Recruitment Procedures

1. If applicable, describe how potential participants will be initially identified (e.g., public records, private medical records, private school records, etc.).

Potential participants have already been identified by the PI of each participating institution. These lists will be shared with me so that I can contact them to participate in the research aspect of the project.

2. Recruitment is the process by which potential participants are informed of, and invited to participate in, the research study. Check all processes that apply below.

   Recruitment Process
   - Electronic media (e.g., listserv, emails, social media); please specify below

3. Describe when, where, and how participants will be initially contacted. Provide specific details as requested above.

The AACR faculty development project will begin with an in-person kick-off meeting. One month prior to this meeting, I will contact potential participants by email.

4. Describe any follow-up recruitment procedures.

I will re-send the email one week and a few days before the kick-off meeting, if some potential participants have not yet responded.

Data Collection Instruments

1. List and describe all the instruments (interview guides, questionnaires, surveys, observation guides, etc.) to be used for this study. Attach a copy of all instruments.

<table>
<thead>
<tr>
<th>Instrument Name</th>
<th>Instrument Description</th>
<th>Groups that will Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wieman Teaching Practices Inventory</td>
<td>This inventory is designed to understand the teaching practices of participants.</td>
<td>All faculty participants.</td>
</tr>
<tr>
<td>Approaches to Teaching Inventory</td>
<td>A valid and reliable instrument that assesses instructors’ overall approach to teaching.</td>
<td>Faculty participants</td>
</tr>
<tr>
<td>Semi-structured interview_faculty</td>
<td>Interview questions to be asked of faculty participants.</td>
<td>Faculty</td>
</tr>
<tr>
<td>Teaching Practices Inventory</td>
<td>This inventory is designed to understand the teaching practices of participants.</td>
<td>all participants</td>
</tr>
<tr>
<td>Administrator interview and survey questions</td>
<td>Potential items that will be used to collect data from administrators. These Administrators items will be refined and revised before data collection begins.</td>
<td>Faculty participants' classrooms will be</td>
</tr>
<tr>
<td>COPUS</td>
<td>This instrument is an observation protocol for classroom observations. Using this form, researchers will log what instructors and students are doing every 2 minutes. They will not be recording any information that provides individual identifiable information about students in the Faculty participants' classrooms will be</td>
<td></td>
</tr>
</tbody>
</table>
classroom. The attachment is a report from the primary literature that shows data on validation of the instrument as well as details about use of the tool.

* Attachments are available under the Documents tab of project home page

View: IRB_Risks_and_Benefits

**Risks and Benefits**

**Important Note:** Do not respond with "Not Applicable" or "N/A" for any of the questions in this section. A complete statement must be provided for every item in this section (for example, if there are no risks or benefits, please state this). *If there are procedures that participants would receive that are not considered 'research' activities (e.g., regular educational practice or standard medical treatment), please limit description of risks and benefits to procedures that will be conducted solely for 'research' purposes.*

1. **Risks and/or Discomforts**

   **Describe any reasonably foreseeable psychological, social, legal, economic or physical risks and/or discomforts from all research procedures, and the corresponding measures to minimize these.**

   **Important Note:** *If there is more than one study procedure or more than one discrete subject group, please identify the procedure or group and separate the descriptions for both (a) and (b).*

   **(a) Risks and/or Discomforts.**

   There are three potential risks associated with participating in this study. First, as with any study that uses identifiers, there is a risk of breach of confidentiality. Second, there is a risk of slight psychological discomfort resulting from discussing individual beliefs, knowledge, experiences, and practices regarding teaching and learning. Teaching at the college level is generally a very solo endeavor, so many participants may not have spoken candidly about these topics before. Third, for interviews held via phone or online conferencing, we cannot guarantee against a breach of confidentiality during the interview.

   **(b) Measures to minimize the risks and discomforts to participants.**

   First, we will minimize the risk of breach of confidentiality by using indirect identifiers to label the data we collect. Only the research team will have access to the key that links the identifiers to names and contact information. This list will be kept in a locked cabinet in a locked office on the UGA campus. The coded data will be kept in electronic data files on a password-protected computer in a locked office. If a participant declines to be contacted for follow-up questions and/or interviews, the transcript and notes will have a code that is not linked to any identifying information. Audio files will be transcribed within two (2) months of data collection, will be retained for three years until the results are published, and will then be destroyed. Second, we will minimize the psychological discomfort experienced by participants by acknowledging the personal nature of the questions before and throughout the interview. We will encourage participants to decline to answer any question that makes them uncomfortable. We will fully explain that the purpose of the research is to focus on the experience of faculty members and that their personal experiences with the instructional change process can help us to understand what support would help facilitate faculty change. We will also limit our interviews to one hour. Third, we will inform participants in phone and online interviews that we cannot guarantee against a breach of confidentiality during the interview.

2. **Benefits**

   **Describe any potential direct benefits to study participants. If none, indicate so.** *Important Note* - *Please do not include compensation/payment/extra credit in this section, as these are "incentives" and not "benefits" of participation in research; any incentives must be described in the "Human Research Participants" section.*
We do not foresee any direct benefits to you as a result of participating in this study.

**Describe the potential benefits to society or humankind.**

College biology instructors are an essential element in education reform, yet we know little about the instructional change process they undergo. This work will help us to understand their experiences and has the potential to inform entities that strive to support faculty, such as professional development programs and administrators.

3. **Risk/Benefit Analysis**

   **Indicate how the risks to the participants are reasonable in relation to the anticipated benefits, if any, to participants and the importance of the knowledge that may reasonably be expected to result from the study. (i.e., How do the benefits of the study outweigh the risks, if not directly to the participants then to society or humankind?).**

   The potential benefits of this work could have an important impact on how instructors are viewed as part of education reform, and the risks to the participants can be greatly reduced by the measures described above.

4. **Sensitive or Illegal Activities**

   **Will the study collect any information that if disclosed could potentially have adverse consequences for participants or damage their financial standing, employability, insurability, or reputation (includes but not limited to sexual attitudes, preferences, or practices; HIV/AIDS or other sexually transmitted diseases; use of alcohol, drugs, or other addictive products; illegal conduct; an individual's psychological well-being or mental health; and genetic information)?**

   ○ Yes  ○ No

   **If "YES", respond to (a) and (b) below. Include in the consent document(s) a description of the information and potential risk or harm to participants from the collection of this information.**

   (a) **Describe specific information that will be collected:**

      N/A

   (b) **Describe how the researchers will protect this information from any inadvertent disclosure:**

      N/A

5. **Reportable Information**

   **Is it reasonably foreseeable that the study will collect or be privy to information that State or Federal law requires to be reported to other officials (e.g., child or elder abuse) or ethically might require action (e.g., suicidal ideation, intent to hurt self or others)?**

   ○ Yes  ○ No

   **If "YES", please explain and include a discussion of the reporting requirements in the consent document(s).**
Data Security

1. **Data Security:**

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidential</td>
<td>The responses/information may potentially be linked/tracked back to an individual participant, for example, by the researchers (like in face to face interviews, focus groups). If necessary, provide additional information below.</td>
</tr>
<tr>
<td>Direct/Indirect Identifiers</td>
<td>The data and/or specimens will be labeled with a direct identifier or an indirect identifier (code) that the research team can link to individually identifiable information but the results, when disseminated, will not be individually identifiable. Provide additional information below.</td>
</tr>
</tbody>
</table>

2. **Confidential - Indirect Identifiers:** The data and/or specimens will be labeled with a code that the research team can link to individually identifiable information using a code key or master list. If the data and/or specimens will be coded, please select one or more options and provide additional information below.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer/electronic files will be used</td>
<td>The key to the code will be in an encrypted and/or password protected file. The coded data file will be maintained on a separate computer/server.</td>
</tr>
</tbody>
</table>

3. **Confidential - Direct Identifiers:** If the data and/or specimens will be labeled with individually identifiable information, please select one or more options and provide additional information.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>There are no items to display</td>
<td></td>
</tr>
</tbody>
</table>

4. **If Confidential-Indirect Identifiers or Confidential-Direct Identifiers are being used, please explain why these are necessary.**

   Indirect identifiers will be used to link multiple pieces of data from the same participants and so that we can follow-up with interview participants if our analysis reveals additional questions or if they agree to participate in additional interviews.

5. **Identify who will have access to the individually identifiable information and/or the key to the code.**

   Only members of the research team will have access to the individually identifiable information and the key to the code.

6. **Provide more information as requested above.**

Future Use - Handling of Information After Collection

1. **If individually identifiable information and/or codes will be retained after completion of data collection, describe how the information will be handled or stored to ensure confidentiality. Check all that apply.**

<table>
<thead>
<tr>
<th>Option</th>
<th>Handling Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifiable information retained</td>
<td>Individually identifiable information and/or codes linking the data or specimens to individual identifiers will be retained. Describe retention period below.</td>
</tr>
<tr>
<td>Audio/Video Retained</td>
<td>Audio and/or video recordings (if applicable) will be retained. Describe retention period below.</td>
</tr>
</tbody>
</table>
2. If direct identifiers, code keys/master lists, and/or audio/video recordings will be retained after data collection is complete, please provide the following additional information.

   **Retention Period:**

   3 years

   **Retention Justification:**

   We will retain interview recordings to link multiple interviews from single participants.

   **Procedure for removing or destroying the direct/indirect identifiers, if applicable:**

3. If "other" is selected, provide additional information below:

**Consent Process**

Important Note: The IRB strongly recommends the use of consent templates that are available on the consent materials page to ensure that all the elements of informed consent are included (per 45 CFR 116). If more than one consent document will be used, please name each accordingly.

1. **Describe how, where, and when informed consent will be obtained from research participants (or permission from parent/s or guardian/s and assent from minor participants). Select the applicable option(s) below and provide a detailed description where prompted.**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informed consent will be obtained and documented</td>
<td>All consent documents that participants will sign have been attached.</td>
</tr>
</tbody>
</table>

2. **Please describe how, where and when informed consent will be obtained.**

   Informed consent will be obtained from participants at the first interview.

3. **The IRB may waive the requirement to obtain a signed consent form for some or all subjects if it finds that the project meets one of the criteria below. Please choose the appropriate criterion and provide supporting justification for requesting the waiver.**

   **Provide supporting justification for requesting a waiver to document informed consent.**

4. **The IRB may approve a consent procedure which does not include, or which alters, some or all of**
the elements of informed consent set forth in 45 CFR 46.116, or waive the requirements to obtain informed consent provided the IRB finds and documents that:

a. The research involves no more than minimal risk to the participants;
b. The waiver or alteration will not adversely affect the rights and welfare of the participants;
c. The research could not practicably be carried out without the waiver or alteration; and,
d. Whenever appropriate, the participants will be provided with additional pertinent information after participation.

Provide a justification for requesting a waiver of informed consent or a waiver or alteration of the elements of informed consent that addresses each of the criteria a-d above.

View: SF: UGA_IRB_Methods_Proc_Req_Add_Info

Methods and Procedures that Require Additional Information

1. Check all that apply. Important Note: The items listed below are NOT an inclusive list of methods and procedures that may be used in research studies. Some procedures require the completion of specific sections on other pages of the submission.

   Method/Procedure: Internet Research

   If "Other" is selected, please provide further information.

View: SF: UGA_IRB_Incomplete_Disclosure

Deception, Concealment, or Incomplete Disclosure

1. Describe the deception, concealment, or incomplete disclosure; explain why it is necessary, and procedure for debriefing the participants. Important Notes: 1. A request for the IRB to waive or alter elements of consent should be requested in the Consent Process section. 2. The consent document should include the following statement: "In order to make this study a valid one, some information about (your participation or the study) will be withheld until completion of the study."

2. Attach Debriefing Form:

3. If Debriefing Form is not attached, please explain why.

View: SF: UGA_IRB_Internet_Research

Internet Research

1. If data will be collected, transmitted, and/or stored via the internet, the level of security should be appropriate to the level of risk. Indicate the measures that will be taken to ensure security of data transmitted over the internet. Check all that apply.

   Security Measure
   A mechanism will be used to strip off the IP addresses for data submitted via e-mail.
   The data will be transmitted in encrypted format.
Firewall technology will be used to protect the research computer from unauthorized access. Hardware storing the data will be accessible only to authorized users with log-in privileges. Other (please describe), or provide additional pertinent information.

2. **If "Other" is selected, please provide additional pertinent information.**

Some interviews may be conducted via online conferencing. For these interviews, there is no way to guarantee against a breach of confidentiality. We will inform participants of this fact.

For online surveys, responses will be stripped of indetifiers such as ip address and email address in order to protect the confidentiality of participants. If supported on the end user's browser, SSL will be setup as the preferred option.

View: IRB_Blood_Sampling

**Blood Sampling/Collection**

If blood will be collected for the purpose of this research, please respond to all the following:

1. **Route/method of collection (e.g., by finger stick, heel stick, venipuncture):**

2. **Frequency of collection (e.g., 2 times per week, for 3 weeks):**

3. **Volume of blood for each collection (in milliliters):**

4. **Are participants healthy, non-pregnant adults who weigh at least 110 pounds? If no, indicate if amount collected will exceed the lesser of 50 ml or 3 ml per kg in an 8-week period and if collection will occur more frequently than 2 times per week.**

5. **Will participants fast prior to blood collection(s)? If yes, describe how informed consent will be obtained prior to fasting.**

View: SF: Supporting Documents

**Supporting Documents**

Attach supporting files, naming them as you want them to appear in the approval letter:

<table>
<thead>
<tr>
<th>Document</th>
<th>Category</th>
<th>Date Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are no items to display

Suggested attachments:

- Conflict of Interest Committee’s determination for any financial interest related to the research
- Complete checklist of meeting Department of Energy requirements, if applicable
- Other study-related documents not attached on previous forms
Goals and Procedures
The goal of this study is to better understand the experiences of college science faculty who participate in Faculty Learning Communities associated with the Automated Analysis of Constructed Response (AACR) project. If you volunteer to take part in this study, you will be asked to:

- Answer questions about your attitudes about the AACR assessments and reports and your perception of the systemic structures that facilitate and impede sustained use of these assessments. You will be interviewed three times over three years, and the interviews will take no more than 1 hour.
- Agree to be audiorecorded during the interviews.
- Respond to the Approaches to Teaching Inventory, which measures your overall approach to teaching. This survey will take about 15 minutes. You will respond to the survey three times over three years in an online format.
- Allow us to use records of your use of the AACR web portal and attendance at Faculty Learning Community meetings as data for research.
- Give the PI permission to email you to clarify your responses and/or to request further information about your teaching, such as syllabi or exams.

Potential Risks or Discomforts
We foresee minimal risks and discomforts resulting from participating in this study. No names or identifying information will be included on materials seen by anyone outside of the research team. All data, including interview transcripts and the interviewers notes, will have a code and only the research team will have access to that code, which will be kept in a locked cabinet in a locked room. Despite these preventative measures, we cannot eliminate all risk of a breach of confidentiality. Additionally, the surveys will be administered online and we may email you following the interviews. Internet communications are insecure and there is a limit to the confidentiality that can be guaranteed due to the technology itself. However, once the materials are received by the researcher, standard confidentiality procedures will be employed.

Since college teaching tends to be such an individual pursuit, some of the questions we ask may elicit ideas you have rarely or never shared with others. This may feel slightly uncomfortable at first, and you can decline to answer any question you do not wish to answer.

Potential Benefits
We do not foresee any direct benefits to you as a result of participating in this study. The benefit of your participation is that you can contribute to research aimed at determining the support instructors need to facilitate instructional change.
Voluntary Participation and Freedom to Withdraw
Your participation is voluntary. You can refuse to participate and you are free to withdraw from this study at any point without penalty or loss of benefits to which you are otherwise entitled. If you decide to withdraw from the study, the information that can be identified as yours will be kept as part of the study and may continue to be analyzed, unless you make a written request to remove, return, or destroy the information.

Confidentiality
Participation in the study is confidential. Data will be anonymized prior to dissemination. This means that no names or identifying information about you or the institution where you work will be included in any materials or publications resulting from this research. The public will not have access to identifiable data collected from you.

Audiorecordings and data files with identifiers will be retained for three (3) years. The data will be on a secure, password-protected computer in a locked room or in a locked filing cabinet. Final findings from this research will be prepared for publication in a peer-reviewed journal, and will not include identifying information. Although your name will never be attached to them, you will be given the opportunity to review direct quotations used for publications or presentations.

The researchers will answer any further questions about the research, now or during the course of the project, and can be reached by telephone at 706-542-9616.

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

Paula P. Lemons
Name of Researcher Signature Date

Telephone: __706-542-9616_________

Email: ___plemons@uga.edu_________

_________________________ Signature Date

Name of Participant Date

Please sign both copies, keep one and return one to the researcher.

Additional questions or problems regarding your rights as a research participant should be addressed to The Chairperson, Institutional Review Board, University of Georgia,
Consent form adapted from Borgatti and Molina (2005)
Study: Experiences of Faculty Participants in the Automated Analysis of Constructed-Response Project
Informed Consent Form: Faculty Form

You are invited to participate in a research study being conducted by Drs. Paula P. Lemons (The University of Georgia, 706-542-9616, plemons@uga.edu), Mary Anne Sydlik (Western Michigan University, 269-387-5393, maryanne.sydlik@wmich.edu), John Merrill (Michigan State University, 517-432-3618, merrill3@msu.edu) Mark Urban-Lurain (Michigan State University, 517-432-2108, urban@msu.edu), Jennifer Knight (University of Colorado, Boulder, 303-735-1949, knight@colorado.edu), Luanna Prevost (University of South Florida, 813-974-7836, prevost@usf.edu), Michelle K. Smith (University of Maine, 207-581-2604, michael.k.smith@maine.edu), Ross Nehm (State University of New York, Stony Brook, 631-632-7247, ross.nehm@stonybrook.edu), Tessa Andrews (The University of Georgia, 706-542-3340, tandrews@uga.edu), and Meghan Federer (The Ohio State University, 614-648-3462, federer.21@osu.edu).

Goals and Procedures
The goal of this study is to better understand the experiences of college science faculty who participate in Faculty Learning Communities associated with the Automated Analysis of Constructed Response (AACR) project. If you volunteer to take part in this study, you will be asked to:
• Allow observation of your classroom several times per semester in order to note how class time is spent. These observations will be arranged in advance with your permission.
• Share your course materials, such as exams, with the research team.
• Complete a survey about your teaching experiences. This survey will take about 20 minutes. You will respond to the survey three times over three years in an online format.
• Participate in interviews about your attitudes about the AACR assessments and reports, teaching, and student learning. You will be interviewed multiple times, and interviews will take no more than 1 hour.
• Agree to be audio recorded during the interviews and Faculty Learning Community meetings.
• Allow the research team to use records of your use of the AACR web portal and attendance at Faculty Learning Community meetings as data for research.
• Give the research team permission to email you to clarify your responses and/or to request further information about your teaching, such as exams.
• Agree to participate in evaluation of the AACR project conducted by Science and Mathematics Program Improvement (SAMI) from Western Michigan University. SAMI will ask you to complete select pre/post surveys and other evaluation instruments, and may ask you to participate in an annual interview/focus group. Interviews/focus groups may be conducted over the phone or through video conferencing. Your participation in SAMI evaluation activities will require about 4 hours annually.
• This is a longitudinal study that lasts three years, which is the scheduled duration of the Faculty Learning Communities.

Potential Risks or Discomforts
We foresee minimal risks and discomforts resulting from participating in this study. No names or identifying information will be included on materials seen by anyone outside of the research team. All data, including interview transcripts and the interviewers notes, will have a code and only the research team will have access to that code, which will be kept in a locked cabinet in a locked room. Despite these preventative measures, we cannot eliminate all risk of a breach of confidentiality. Additionally, the surveys will be administered online, and we may email you following the interviews. Internet communications are insecure and there is a limit to the confidentiality that can be guaranteed due to the technology itself. However, once the research team receives the materials, standard confidentiality procedures will be employed.

Since college teaching tends to be such an individual pursuit, some of the questions we ask may elicit ideas you have rarely or never shared with others. This may feel slightly uncomfortable at first, and you can decline to answer any question you do not wish to answer.

Potential Benefits
We do not foresee any direct benefits to you as a result of participating in this study. The benefit of your participation is that you can contribute to research aimed at determining the support instructors need to facilitate instructional change.
Goals and Procedures
The goal of this study is to better understand the attitudes of college administrators who oversee faculty who participate in Faculty Learning Communities associated with the Automated Analysis of Constructed Response (AACR) project. If you volunteer to take part in this study, you will be asked to:

• Answer questions about your attitudes about members of your faculty who participate in the AACR project. You will be interviewed three times over three years, and the interviews will take no more than 1 hour.
• Agree to be audiorecorded during the interviews.
• Respond to a survey about your views on faculty participation in the AACR project. The survey will take about 15 minutes. You will respond to the survey three times over three years in an online format.
• Give the PI permission to email you to clarify your responses and/or to request further information.

Potential Risks or Discomforts
We foresee minimal risks and discomforts resulting from participating in this study. No names or identifying information will be included on materials seen by anyone outside of the research team. All data, including interview transcripts and the interviewers notes, will have a code and only the research team will have access to that code, which will be kept in a locked cabinet in a locked room. Despite these preventative measures, we cannot eliminate all risk of a breach of confidentiality. Additionally, the surveys will be administered online and we may email you following the interviews. Internet communications are insecure and there is a limit to the confidentiality that can be guaranteed due to the technology itself. However, once the materials are received by the researcher, standard confidentiality procedures will be employed.

Potential Benefits
We do not foresee any direct benefits to you as a result of participating in this study. The benefit of your participation is that you can contribute to research aimed at determining the support instructors need to facilitate instructional change.

Voluntary Participation and Freedom to Withdraw
Your participation is voluntary. You can refuse to participate and you are free to withdraw from this study at any point without penalty or loss of benefits to which you are otherwise entitled. If you decide to withdraw from the study, the information that can be identified as yours will be kept as part of the study and may continue to be analyzed, unless you make a written request to remove, return, or destroy the information.
**Confidentiality**

Participation in the study is confidential. Data will be anonymized prior to dissemination. This means that no names or identifying information about you or the institution where you work will be included in any materials or publications resulting from this research. The public will not have access to identifiable data collected from you.

Audiorecordings and data files with identifiers will be retained for three (3) years. The data will be on a secure, password-protected computer in a locked room or in a locked filing cabinet. Final findings from this research will be prepared for publication in a peer-reviewed journal, and will not include identifying information. Although your name will never be attached to them, you will be given the opportunity to review direct quotations used for publications or presentations.

The researchers will answer any further questions about the research, now or during the course of the project, and can be reached by telephone at 706-542-9616.

*I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.*

Paula P. Lemons
Name of Researcher ____________________ Signature ____________ Date ____________

Telephone: __706-542-9616___________

Email: ___plemons@uga.edu___________

__________________________
Name of Participant Signature ____________ Date ____________

Please sign both copies, keep one and return one to the researcher.

Additional questions or problems regarding your rights as a research participant should be addressed to The Chairperson, Institutional Review Board, University of Georgia, 629 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-3199; E-Mail Address IRB@uga.edu

Consent form adapted from Borgatti and Molina (2005)
Dear AACR Faculty Learning Community Participant,

I am a researcher from the University of Georgia, and I am investigating the experiences of college science faculty who participate in Faculty Learning Communities associated with the Automated Analysis of Constructed Response (AACR) project.

If, in addition to your participation in the FLC, you are willing to participate in the research described below, please respond to this email and include the information requested at the end of the message.

If you agree to participate in this research, I will collect data from you throughout participation in the FLC. More specifically, I will ask you to

- Answer questions about your attitudes about the AACR assessments and reports and your perception of the systemic structures that facilitate and impede sustained use of these assessments. You will be interviewed three times over three years, and the interviews will take no more than 1 hour.
- Agree to be audiorecorded during the interviews.
- Respond to the Approaches to Teaching Inventory, which measures your overall approach to teaching. You will respond to the survey three times over three years in an online format.
- Allow the research team to use records of your use of the AACR web portal and attendance at Faculty Learning Community meetings as data for research.
- Give me permission to email you to clarify your responses and/or to request further information about your teaching, such as syllabi or exams.

The first set of interviews will be conducted at the kick-off meeting. Subsequent interviews will be conducted by phone or online conferencing in Year 2 and at the closing meeting of the project in Year 3. All interviews will be conducted at a mutually agreeable time.

Although there are no direct benefits to you for participating in the study, the experiences and information you provide will be useful for improving our understanding of the support college instructors need to successfully implement new strategies that interest them. Gathering responses from as many participants as possible will enable us to represent the experiences of a diverse group of faculty and institutions.

Thank you for considering participating. If you are willing to participate, please email us at plemons@uga.edu and include the following information:

1. Your name:
2. Your institution:

Please feel free to contact me if you have any questions or would like additional information. Thank you for considering this invitation.
Best wishes,

Paula Lemons, Assistant Professor
Department of Biochemistry and Molecular Biology
University of Georgia
plemons@uga.edu
Dear Administrator,

I am a researcher from the University of Georgia, and I am investigating the experiences of college science faculty who participate in Faculty Learning Communities associated with the Automated Analysis of Constructed Response (AACR) project.

I am contacting you because I am interested in learning, not only about the experiences of faculty participants in AACR, but also about the attitudes of administrators who oversee them.

If you agree to participate in this research, I will collect data from you throughout the duration of your faculty members’ participation in the FLC. More specifically, I will ask you to

- Answer questions about your attitudes about members of your faculty who participate in the AACR project. You will be interviewed three times over three years, and the interviews will take no more than 1 hour.
- Agree to be audiorecorded during the interviews.
- Respond to a survey about your views on faculty participation in the AACR project. You will respond to the survey three times over three years in an online format.
- Give the PI permission to email you to clarify your responses and/or to request further information.

If possible, interviews will be conducted face-to-face, but if time and distance prevents face-to-face interviews, I will interview you by phone or online conferencing. All interviews will be conducted at a mutually agreeable time.

Although there are no direct benefits to you for participating in the study, the experiences and information you provide will be useful for improving our understanding of the support college instructors need to successfully implement new strategies that interest them. Gathering responses from as many administrators as possible will enable us to represent the experiences of a diverse group of administrators and institutions.

Thank you for considering participating. If you are willing to participate, please email us at plemons@uga.edu and include the following information:

1. Your name:
2. Your institution:
3. Your administrative role at your institution:

Please feel free to contact me if you have any questions or would like additional information. Thank you for considering this invitation.

Best wishes,

Paula Lemons, Assistant Professor
Department of Biochemistry and Molecular Biology
University of Georgia
plemons@uga.edu
The Teaching Practices Inventory: A New Tool for Characterizing College and University Teaching in Mathematics and Science

Carl Wieman* and Sarah Gilbert†

*Department of Physics and Graduate School of Education, Stanford University, Stanford, CA 94305; †Carl Wieman Science Education Initiative, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Submitted February 8, 2014; Revised June 7, 2014; Accepted June 7, 2014
Monitoring Editor: Erin Dolan

We have created an inventory to characterize the teaching practices used in science and mathematics courses. This inventory can aid instructors and departments in reflecting on their teaching. It has been tested with several hundred university instructors and courses from mathematics and four science disciplines. Most instructors complete the inventory in 10 min or less, and the results allow meaningful comparisons of the teaching used for the different courses and instructors within a department and across different departments. We also show how the inventory results can be used to gauge the extent of use of research-based teaching practices, and we illustrate this with the inventory results for five departments. These results show the high degree of discrimination provided by the inventory, as well as its effectiveness in tracking the increase in the use of research-based teaching practices.

INTRODUCTION

Research has shown the effectiveness of particular teaching practices in science, technology, engineering, and mathematics (STEM), such as more active and collaborative learning. There have been many calls for the greater adoption of such research-based teaching practices, originating from, among others, the National Research Council (NRC, 2012), the President’s Council of Advisors on Science and Technology (PCAST, 2012), and the Association of American Universities (AAU, 2011).

A major difficulty in achieving the desired change is that the teaching practices used in college and university STEM courses remain largely unmeasured. At the request of one of us (C.W.) the AAU and the American Public and Land Grant Universities polled their members on whether or not they collected data on the teaching practices used in their STEM courses. C.W. also posed the same question to the attendees of the annual meeting of the Presidents and Chancellors of the Association of American Colleges and Universities. No institution reported collecting data on the teaching practices in use in its courses.

To our knowledge, no method currently exists for collecting such data in an efficient and consistent manner. The only data on teaching collected at most universities (Berk, 2005) are student course evaluations, but these provide little information on the teaching practices and little guidance to instructors as to how to improve (Cohen, 1980). There are a number of classroom observation protocols for undergraduate STEM that have been developed and validated, such as the Reformed Teaching Observation Protocol (Sawada et al., 2002), the Teaching Dimensions Observation Protocol (Hora et al., 2013), and the Classroom Observation Protocol for Undergraduate STEM (COPUS; Smith et al., 2013). While all of these provide useful data, classroom observation protocols necessarily capture only the classroom elements of the practices that go into teaching a course. They also require hours of training and observations to adequately characterize this fraction, as classroom activities can vary from one day to the next.

DOI: 10.1187/cbe.14-02-0023
Address correspondence to: Carl Wieman (cwieman@stanford.edu).
© 2014 C. Wieman and S. Gilbert. CBE—Life Sciences Education © 2014 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/licenses/by-nc-sa/3.0).
“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.
The teaching practices inventory (TPI) presented in this paper is designed to allow the broader range of practices that are involved in teaching a STEM course to be quickly determined. As such, it is possible to use that information to then determine the extent of use of research-based practices. To facilitate that determination, we have created a scoring rubric that extracts a numerical score reflecting the extent of use of research-based practices. Use of the inventory helps instructors evaluate their teaching, see how it might be improved, and track improvement.

The PULSE Vision and Change course-level rubric (PULSE, 2013) is in a similar spirit to our TPI and scoring rubric. All seven factors listed in that PULSE rubric can be seen to be reflected in items on the TPI. However, the TPI is designed to provide a more extensive and detailed characterization of the teaching in each individual course.

DEVELOPMENT AND VALIDATION

The full 72-item inventory with scoring rubric is given in the Supplemental Material, but we provide a few items here as typical examples. (On the actual inventory there are check boxes that are filled out to indicate whether a listed practice is used in the course or provided to the students.)

Assignments with feedback before grading or with opportunity to redo work to improve grade
Students see marked assignments
List of topics to be covered
List of topic-specific competencies (skills, expertise ... students should achieve (what students should be able to do)
Assessment given at beginning of course to assess background knowledge
Teaching assistants receive one-half day or more of training in teaching

The items on the inventory are divided into eight categories, as shown in Table 1.

We are using the term “inventory” in its conventional meaning of a list of all items present, in this case a list of all the teaching practices present in a course. This is different from the meaning the word “inventory” has taken on in a science education research context, namely an instrument for the measurement of mastery of some particular scientific concept, such as the Genetics Concept Assessment (Smith et al., 2008) or the Force Concepts Inventory (Hestenes, 1992). This difference has implications for the development and validation of the instrument. The “construct” to be measured in this case is the set of teaching practices that are commonly or occasionally used in math and science courses. Our definition of “occasional” (as distinguished from very infrequent or unique) is that, to our knowledge, the practice has been used in multiple science or mathematics courses distributed across four or more different universities or colleges. To be valid as an inventory, the TPI has to accurately characterize the range of teaching practices used in a course when an instructor makes a good faith effort to complete the inventory. Our primary testing and refinement focused on ensuring that science and math instructors will interpret the items in a consistent and accurate manner and that the inventory covered all teaching practices used by more than two instructors in our large test sample. Owing to the nature of this construct, the statistical tests that one would use to check reliability and validity of a conventional instrument like the genetics concept assessment are not applicable in this case. In particular, tests of the relationships between items do not provide meaningful information about the quality of the assessment instrument. Finally, the inventory only tells whether a practice is being used, it does not tell the quality of implementation. As discussed in the Further Work section, we have some evidence that it is far more difficult to measure quality of implementation of practices.

The development process involved two major iterations and one final round of minor revisions. The first iteration was in 2007. At that time, we were trying to characterize the teaching practices in use in the science departments at the University of British Columbia (UBC) at the launch of the Carl Wieman Science Education Initiative (CWSEI). The instructors in math and sciences at UBC are quite similar to...
the instructors at any large U.S. public research university. A substantial fraction are from the United States, and most of them have either studied or taught at U.S. universities at some point in their careers. We developed the inventory relatively quickly, relying on our own knowledge of the education research literature and our experience with science instructors and faculty development across several science departments while working on the University of Colorado Science Education Initiative (CU-SEI). We shared a draft of the inventory with about a dozen instructors in the UBC science departments, and their feedback was used to refine the wording to improve the clarity. Approximately 150 instructors then completed that first version of the inventory.

Over the next several years, we created a second version, guided by the 150 responses and associated feedback on the first version and from the extensive experience gained on instructors’ teaching practices through the work of the CWSEI. In developing the second version, we examined all the inventory responses to see where there was evidence of confusion over the questions, where there were frequent responses in the “other” categories (therefore not one of the listed response options), or whether some items seemed unnecessary or inappropriate. We also analyzed all of the open-ended comments from the instructors. These were easily coded into categories of: 1) said they were using a practice marked as “other” that matched something we had intended to cover by one of the response options, thus indicating confusion as to the description of the options; 2) described a practice relevant to that item but not covered by any listed option; or 3) described a practice they could not see how to capture using any of the items on the TPI. The total number of comments in all three categories for any item was well below 10% of the total responses for that item, indicating there were no serious problems with the version 1 items.

However, there were a number of places where it was possible to make minor improvements. We also added a few items and response options to capture a larger range of practices, as determined from the combination of: the review of the version 1 inventory responses, informal discussions with many instructors across the departments of mathematics and the sciences, and systematic review of the inventory and input on practices observed from the ∼30 science education specialists (SEs; Wieman et al., 2010) who worked with a number of different instructors in all of the math and science departments during that period of time. The SEs were able to provide full descriptions of the teaching practices used by nearly all of the instructors in three departments with large CWSEI programs, as well as descriptions of the practices used by a substantial fraction of the instructors working in other departments affiliated with CWSEI and CU-SEI. We also added items on teaching assistant (TA) selection, training, and guidance, and made a number of minor wording changes to improve clarity.

We organized the questions into the eight categories listed based on usability interviews and feedback. These categories and the format of the survey were chosen only to make completion of the survey easier, not for any theoretical reason, and were finalized only after we had determined all the practices/items we wanted to include. Feedback from the SEs and CWSEI department directors and discussions with other instructors indicated these categories tended to match how instructors generally organized their thinking about the different elements of teaching a course, and so this organization made the process of filling out the survey most efficient.

After completing these revisions, we had three other experts in college science teaching2 and the SEs review this draft of the second version of the inventory. They made suggestions for minor changes, most of which were incorporated.

Finally, the five instructors who served as the CWSEI departmental directors in the science and math departments, and hence are representatives from each discipline, carefully went over each question as the final stage of the second iteration. They filled out the inventory for the courses they were teaching and then went through their responses and interpretations of the questions with us. We assessed whether they interpreted what was being asked as we intended and elicited their opinions as to whether the instructors in their departments might find any question confusing or misleading. This process led to a few more very minor wording modifications, resulting finally in version 2 of the inventory. In spite of this extensive review, 80% of the 2007 items ended up either unchanged or with only very slight changes in wording in version 2.

To improve the accuracy and consistency of responses, we designed the inventory to minimize the number of subjective judgments. Only two items are likely to have substantial subjectivity in the responses, and these are both in category III: in-class features and activities. These are the items: “How many times do you pause to ask for questions [during class]?” and “Fraction of typical class time spent lecturing?” We particularly recognized the limitations of the first question but decided to keep it, because it is meaningful, and there is value to encouraging instructor reflection on this specific item. From our experience, we expected that the estimates of fraction of time spent lecturing would be more clearly defined in the minds of the instructors and the responses more accurate than for the first question, but still rather subjective. As discussed in the Accuracy of Responses section, we have conducted some testing of the “fraction of typical class spent lecturing” responses.

During the development process, we discovered that the formats and instructional practices of courses labeled as “labs” (including project-based) and “seminar courses” (where the structure and pace of the class was largely driven by students, rather than an instructor) were highly idiosyncratic and varied widely from one course to the next. We were unable to find meaningful common features by which to characterize the teaching practices used in such courses, and so we recommend that the TPI not be used with them. The educational goals also varied widely across all such courses that we analyzed and were usually ill defined, making it difficult to determine whether any of the practices used had research evidence indicating their effectiveness. Our observations matched the findings of the NRC review of the research on instructional labs in science (NRC, 2006).

One hundred and seventy-nine instructors from five math and science departments completed version 2 of the inventory. We reviewed all of those responses, particularly all

2Peter Lepage, cochair of the PCAST subcommittee on undergraduate STEM education; Susan Singer, chair of the NRC study of discipline-based education research in science and engineering; and Michelle Smith, biology education researcher and a member of the University of Maine Center for Research in STEM Education.
the responses in the “other” categories and the open-ended responses, looking for indications that the instructors had misinterpreted a question or that they felt they were using practices not captured adequately by the inventory (which also could be the result of misinterpretation). There were only isolated examples of individual instructors misinterpreting an item or a response option. On the third items for which the latter occurred three to five times, we made small wording changes. There were only three instructors who said it was difficult to adequately describe the practices in their courses with the TPI options. We discovered that two of the respective courses were seminar courses and the other was a project lab course. Those three instructors had simply overlooked the instructions telling instructors not to fill out the TPI for courses of those types. Finally, it appeared that three (1.5%) of the instructors gave numbers based on “per term,” rather than “per class” as stated in the item. The primary difference between version 2 and the final version we present in this paper were changes in wording to give greater emphasis to that distinction. The final version of the inventory is given in the Supplemental Material.

**ACCURACY OF RESPONSES**

Our primary validation effort focused on ensuring that the inventory items were interpreted clearly and consistently by instructors and that the inventory captured the practices used by the instructors who completed it. No practices were identified that were used by more than two (of 179) instructors and not captured by the survey. The item interpretation was tested by the department director interviews and the review of the 179 instructor responses. Our assumptions are that when 1) there are no stakes tied to the results, 2) instructors clearly understand what is being asked, and 3) little subjective judgment is required for the response, the responses will likely be accurate. As noted, the latter is true for nearly all of the items in seven of the categories and many of the items in the eighth.

However, we also carried out some limited tests of the accuracy of the responses. The first of these involved having a person other than the instructor check a sample of the responses. Although we recommend having instructors complete the TPI themselves, as there is value to that reflection and it takes the least time, the TPI is not inherently a self-reporting instrument. In most cases, it is easy for another person to determine the correct responses by looking at course materials and instructor class notes. It is more difficult for an independent observer to complete some items of category III: in-class features and activities, as it would require substantial class observation time.

We have selected approximately a dozen random TPI course results and asked the respective SESs in the departments if they thought they were accurate. The SESs are quite familiar with the teaching practices of most of the instructors in their departments. For all but a few cases, they felt they were sufficiently familiar with the instructor and course (or with some review of the course material) to be able to evaluate the accuracy of the responses, and in all those cases, they said they believed the TPI responses were correct to within the width of the levels on the scoring rubric discussed in the Scoring Rubric section, except for the category III items discussed previously.

We also checked with the SESs or CWSEI department directors about several courses that had a surprisingly high or low number of research-based practices. Although we did not get item-by-item evaluation, they confirmed that the general results were reasonable for those instructors according to their knowledge of the teaching practices favored by those instructors.

We compared the TPI responses for seven team-taught courses in which two instructors provided responses for the same course. In five of the team-taught courses, the differences between the TPI responses for different instructors were small (0–2 points using the scoring rubric discussed in the Scoring Rubric section) and consistent with the known differences in classroom practices between the instructors. In two cases, instructors who were team-teaching but were only involved in isolated portions of a course were unaware of some aspects, such as what was provided to students at the beginning of the course, and gave correspondingly inaccurate responses. On the basis of this observation, we believe that, if a course is team-taught, it is best to get a single TPI response from the instructor who is most responsible for the course as a whole. Examining the anomalies also revealed two cases in which “per term” and “per class” labels were apparently misread, as previously noted.

Category III: in-class features and activities is the most difficult for instructors to remember accurately and the most difficult for a third party to check the accuracy of the instructor-supplied TPI data. To address concerns about the accuracy of the TPI responses for category III, we developed an easy-to-use classroom observation protocol, COPUS. This provides a straightforward and efficient way to capture what the instructor and the students are doing during class (CWSEI, 2013; Smith et al., 2013). We have examined the correlation between single-class COPUS observations and instructors’ 2012 TPI responses for 49 courses. Because these were only single-class observations, the results are necessarily crude with respect to any given course, but they did allow us to test whether there were any substantial systematic differences; for example, whether instructors consistently underestimated on the TPI the fraction of time they spent lecturing. We found no systematic differences. The “fraction of class time spent lecturing” for both measures ranges from 10 to 100% for the different courses, and the average overall for the 49 courses is 57% (SD 24%) from the TPI and 58% (SD 28%) from the COPUS observations. There are 16 courses in which the COPUS fraction on the day observed was more than 20% higher than the TPI-reported average fraction of time spent in lecture during the entire term, and 15 courses in which the COPUS observation fraction was more than 20% lower than the TPI value. It is not surprising that the agreement in any particular course is modest, since the TPI is the estimate over an entire term, while the COPUS observations provided a measurement for only a single class period. From multiple COPUS observations of a single course, we know that it is not unusual to have substantial variations from one class to another. This 49-class COPUS sample was from a department in which the fraction of time spent lecturing is relatively low. There are other departments for which a much larger fraction of the TPI responses say that 90–100% of the class time is spent in lecturing. We have limited COPUS data on such higher-lecture-fraction courses, but those data do agree more closely with the TPI data.
We also examined whether overall trends we knew about from other data were accurately reflected in the TPI results. 1) We examined several courses in each department for which we knew there had been substantial efforts supported by the CWSEI to implement research-based instructional practices. The TPI data for those courses reflected those practices and indicated more, usually much more, extensive use of research-based practices than the departmental average. 2) We have a variety of independent measures indicating that the department labeled as D5 in the figures and tables was using fewer research-based practices than other departments, and this was also seen in the TPI results. 3) Finally, we have data indicating that appreciably more than half of the instructors in the department labeled as D3 below have been involved in implementing research-based practices in their teaching in the last several years. The TPI results from 2012–2013 for D3 show significantly greater use of research-based practices than in 2006–2007. These differences are quantified in the Results section.

SCORING RUBRIC
The inventory results in raw form provide an enormous amount of information about how an individual course is taught and, when aggregated by department, about the teaching practices in use in a department. However, it is difficult to quickly determine from the raw inventory results the extent and type of use of research-based practices. To facilitate this determination, we have created a scoring rubric that extracts from the inventory data for each course an “extent of use of research-based teaching practices (ETP)” score for each of the eight inventory categories and for the course as a whole. This rubric assigns points to each practice for which there is evidence of the practice improving learning. The ETP score provides an efficient way to sort through the mass of data provided by the full inventory to identify areas of interest, but it would be a mistake to look at only the ETP score for a course. The breakdown by category and the full inventory response provides a much richer characterization of the teaching.

The first source of evidence used in creating this rubric is the extensive research over the past few decades demonstrating new and more effective teaching practices in science and engineering courses at colleges and universities. These practices have been shown to transcend the specific disciplines and achieve substantially better student learning and other outcomes than the traditional lecture method across the fields of science and engineering (Freeman et al., 2014). These practices are well-known in biology, with evidence of their effectiveness demonstrated in many articles in Science Education and other journals. Examples of such research-based practices are: the use of clicker questions with peer discussion; small-group activities of various types; the use of prereadings with follow-up questions; graded homework; and frequent low-stakes testing and feedback. The National Academy study of discipline-based education research (NRC, 2012) provides the most extensive and authoritative review of this research on the teaching of science and engineering. A new meta-analysis (Freeman et al., 2014) shows gains in both student achievement and course completion that are comparable across the different disciplines. There is also evidence that the amount of student learning that an individual instructor achieves changes when he or she changes the teaching practices he or she is using (Hake, 1998; Knight and Wood, 2005; Derting and Ebert-May, 2010; Hoellwarth and Moelter, 2011; Porter et al., 2013).

The large observed differences in the effectiveness of different science teaching practices and the similarity of those differences across disciplines (Freeman et al., 2014) can be explained in terms of the basic principles of complex learning that have been established by the learning sciences (Bransford et al., 2000; Ambrose et al., 2010; Wieman, 2012). These principles include such things as the need for intense prolonged practice of the cognitive skills desired, with guiding feedback, and the importance of motivation and addressing prior knowledge of the learner. The general learning sciences research is the second source of research literature that was used in creating the scoring rubric. The existence of these underlying principles also implies that it is likely that the relative effectiveness of various teaching practices will also hold for subjects and students for which there are not yet data.

The ideal scoring rubric would assign to each practice a number of points based on a quantitative analysis of its relative benefit to student learning. However, such an analysis to determine the precise weighting would require far more data than currently exist. A much simpler option is to use a binary rubric that merely gives one point to every practice for which there is solid evidence or strong arguments that it supports learning and zero to the rest. We present here a third alternative rubric that is in the spirit of both the Froyd (2008) ranking of promising practices and the PULSE Vision and Change rubrics, wherein they assign broad numerical levels based on qualitative plausibility arguments that are in turn based on the available data, rather than quantitative criteria. Our scoring rubric assigns at least one point to each practice for which there is evidence it supports learning and two or three points to a few practices for which there is evidence suggesting they provide particularly large and robust benefits.

We believe that this rubric provides a more accurate measure of the respective benefits of the teaching practices than a simple binary rubric, but we leave it to the reader to choose which rubric he or she prefers. In either case, a simple Excel spreadsheet can be used to automate the scoring. As shown in the comparison of existing courses and departments found in the discussion of the scoring rubrics in the Supplemental Material, both rubrics provide similar results. When there is more extensive use of research-based practices, it is likely that the differences between rubrics will become more apparent.

The distribution of points for the rubric is shown on the inventory in Appendix 1. The number of points (1–3) given per research-based item depends on our informed but subjective judgments on the consistency, extent, and size of the benefits in the published literature and, to a lesser extent, our experience with the robustness of the benefit from observing (often via the SEs) the adoption of various research-based practices by a few hundred science instructors at the Universities of Colorado and British Columbia. Points are given for a few items, discussed below in this section, for which there is little or no direct published evidence but strong plausibility arguments combined with our observations of instructors’ behaviors and results. We had the same three experts on undergraduate STEM teaching who reviewed the inventory also review the scoring rubric, and they all agreed that it was appropriate.
In Table 2, we provide abbreviated descriptions of all of the inventory items that receive points in the scoring rubric, along with references to the supporting research. The items are grouped according to the nature of their contributions to supporting learning in order to make the comparison between items and supporting research literature more convenient. This categorization is necessarily imperfect, in that a specific practice will often contribute to learning in more than one way and there is some overlap between the listed factors of contributions to learning. We have listed some of the additional contribution types of an item in the table.

The references given in Table 2 to support the scoring are mostly reviews of the literature, rather than specific research studies, since there are an enormous number of the latter for the additional contribution types of an item in the table.

One scoring item that is anomalous is the awarding of 3 points if there are no TAs in the course. These points are not because we feel there is any inherent educational benefit to not having TAs. It is simply to normalize the scoring to make it equivalent for courses that do and do not use TAs. If a course has no TAs, the potential lack of coordination (including coordination of pedagogy) of the TA and non- TA elements of the course and the problems with language fluency of the TAs are not issues, and so an equivalent number of points (3) is provided to courses without TAs.

A common first impression is that this is an excessive set of practices, and that it would not be beneficial to have nearly so many in a course. However, this impression is misleading. First, there are many specific elements involved when you consider all aspects of a course in detail, particularly as it progresses over an entire term. Second, of the 51 items that we have identified as supporting student learning, many of them are used routinely. Third, most of these items are mutually reinforcing, rather than competing for student time and attention. For example, homework and feedback/grading of homework are two elements common to many science courses. The inventory has seven different items relating to how the assignments are given and graded to capture beneficial details, but these do not represent additional activities by the students; they are simply necessary to capture the range of practices used by different instructors. Even though many instructors would have a consistent set of responses across some items, such as the homework and grading choices, it is important to not combine the items, because not all instructors are consistent. Even if there may be substantial correlation between particular item responses when looking at the responses of many instructors together, those differences manifested by some instructors can have significant implications with regard to student learning. Similarly, there are seven items listed under supporting material that we list as beneficial to learning, but most of these will be used by students only occasionally during the course for specific needs.

There are only a few items on the inventory that could conflict, in the sense that they compete for student time and attention if done together. These are all in category III: in-class features and activities. In our opinion, it would not be desirable for an individual course to include frequent ongoing use of every item in category III that we have listed as beneficial, as this would likely be overwhelming for both instructor and student. For nearly all the items in the other categories, good arguments can be made that adding that practice to a course would provide the benefits indicated by the research, without any downsides due to conflicts, assuming the instructor has the time to implement them all adequately.

For the convenience of those who may wish to use the inventory, a clean copy, uncluttered with the scoring and footnotes, is posted at www.cwsei.ubc.ca/resources/TeachingPracticesInventory.htm. An Excel file with formulas to facilitate automatic scoring of responses with the rubric is also available at that website, as is a file of the inventory that can be used to collect inventory data using the Qualtrics online survey tool.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Practice that supports</th>
<th>References on benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Knowledge organization</strong></td>
<td>I. List of topics to be covered</td>
<td>Promising Practice No. 1: Learning Outcomes in Froyd (2008); Chapters 2 and 5 in Ambrose et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>I. List of topic-specific competencies</td>
<td>Promising Practice No. 4: Scenario-based Content Organization in Froyd (2008)</td>
</tr>
<tr>
<td></td>
<td>(+ practice + feedback + metacognition)</td>
<td>Kiewra (1985)</td>
</tr>
<tr>
<td></td>
<td>I. List of competencies that are not topic related</td>
<td>2 Abd-El-Khalick and Lederman (2000)</td>
</tr>
<tr>
<td></td>
<td>(critical thinking, problem solving)</td>
<td>Walberg et al. (1985); Cooper et al. (2006). The reviews by Walberg et al. (1985) and Cooper et al. (2006) are of the extensive K–12 research literature on the beneficial effects of graded homework. Numerous research articles report the educational benefits in undergraduate math and science. Two examples are Cheng et al. (2004) and Richards-Babb et al. (2011).</td>
</tr>
<tr>
<td></td>
<td>II. Animations, video clips, simulations</td>
<td>Kuh (2008)</td>
</tr>
<tr>
<td></td>
<td>II. Lecture notes or copy of class materials1</td>
<td>3 Kuh (2008)</td>
</tr>
<tr>
<td></td>
<td>(partial/skeletal or complete)</td>
<td>Walberg et al. (1985); Cooper et al. (2006). The reviews by Walberg et al. (1985) and Cooper et al. (2006) are of the extensive K–12 research literature on the beneficial effects of graded homework. Numerous research articles report the educational benefits in undergraduate math and science. Two examples are Cheng et al. (2004) and Richards-Babb et al. (2011).</td>
</tr>
<tr>
<td>Long-term memory and reducing cognitive load</td>
<td>II. Worked examples1</td>
<td>Atkinson et al. (2000). Also implies that preclass reading would reduce cognitive load and thereby enhance in-class activities.</td>
</tr>
<tr>
<td></td>
<td>III. Students read/view material on upcoming class and quizzed2</td>
<td>Roediger et al. (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Novak et al. (1999)</td>
</tr>
<tr>
<td>Motivation</td>
<td>I. Affective goals—changing students’ attitudes and perceptions</td>
<td>Chapter 3 in Ambrose et al. (2010); Pintrich (2003); Promising Practice No. 4: Scenario-based Content Organization in Froyd (2008)</td>
</tr>
<tr>
<td></td>
<td>II. Articles from scientific literature</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III. Discussions on why material useful</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V. Students explicitly encouraged to meet individually with you (+ feedback)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VI. Students provided with opportunities to have some control over their learning</td>
<td></td>
</tr>
<tr>
<td>Practice</td>
<td>II. Practice or previous years’ exams + feedback for all items below</td>
<td>Chapter 5 in Ambrose et al. (2010); Promising Practice No. 6: Designing In-class Activities to Actively Engage Students in Froyd (2008); Freeman et al. (2014); Ericsson (2006)</td>
</tr>
<tr>
<td></td>
<td>III. Number of small-group discussions or problem solving</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III. Demonstrations in which students first predict behavior1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III. Student presentations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III. Fraction of class time [not] lecturing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III. Number of PRS questions posed followed by student–student discussion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV. Problem sets/homework assigned and contributing to course grade2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV. Paper or project (involving some degree of student control) (+ knowledge organization + motivation)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V. Fraction of exam mark from questions that require reasoning explanation (+ metacognition)</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>II. Student wikis or discussion board with significant contribution from instructor/TA</td>
<td>Black and William (1998); Hattie and Timperley (2007); Promising Practice No. 5: Providing Students Feedback through Systematic Formative Assessment in Froyd (2008); Chapter 5 in Ambrose et al. (2010); Gibbs and Simpson (2005)</td>
</tr>
<tr>
<td></td>
<td>II. Solutions to homework assignments</td>
<td>Atkinson et al. (2000)</td>
</tr>
<tr>
<td></td>
<td>III. Number of times pause to ask for questions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV. Assignments with feedback and opportunity to redo work (+ metacognition)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV. Students see marked assignments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV. Students see assignment answer key and/or marking rubric</td>
<td></td>
</tr>
<tr>
<td>Metacognition</td>
<td>IV. Students see marked midterm exams</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV. Students see midterm answer keys</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V. Number of midterm exams</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V. Breakdown of course mark</td>
<td></td>
</tr>
<tr>
<td></td>
<td>III. Reflective activity at end of class</td>
<td>Pascarella and Terenzini (2005); Froyd (2008)</td>
</tr>
<tr>
<td></td>
<td>VI. Opportunities for self-evaluation</td>
<td>Chapter 7 in Ambrose et al. (2010); Chapter 3 in Bransford et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>Also all group learning</td>
<td></td>
</tr>
</tbody>
</table>

Continued
Table 2. Continued

<table>
<thead>
<tr>
<th>Factor</th>
<th>Practice that supports</th>
<th>References on benefits</th>
</tr>
</thead>
</table>
| Group learning (has elements of most other categories) | IV. Encouragement for students to work collaboratively on their assignments  
IV. Explicit group assignments  
Also all in-class student discussions | Promising Practice No. 2: Organize Students in Small Groups in Froyd (2008); Chapter 5 in Ambrose et al. (2010) |
| Section 2. Practices that support teacher effectiveness | VI. Assessment at beginning of course  
VI. Use of pre–post survey of student interest and/or perceptions (also feedback on effectiveness) | Bransford et al. (2000); Chapter 1 in Ambrose et al. (2010) |
| Feedback on effectiveness | V. Midterm course evaluation¹  
VI. Use of pre–post survey of student interest and perceptions  
VI. Use of instructor-independent pre–post test (e.g., concept inventory)  
VI. Use of a consistent measure of learning that is repeated  
VI. New teaching methods with measurements of impact on learning | Ericsson (2006) and the other general references above on value of feedback for developing expertise apply here as well.  
¹ Centra (1973); Cohen (1980); Diamond (2004) |
| Gain relevant knowledge and skills | VII. TAs satisfy English-language criteria¹  
VII. TAs receive one-half day or more of training²  
VII. Instructor–TA meetings on student learning and difficulties, etc.²  
VIII. Used “departmental” course materials  
VIII. Discussed how to teach the course with colleague(s)³  
VIII. Read literature about teaching and learning relevant to this course (+ connect with student prior knowledge and beliefs)  
VIII. Sit in on colleague’s class³ | ¹Hinofotis and Bailey (1981); Anderson-Hsieh and Koehler (1988); Jacobs and Friedman (1988); Williams (1992)  
²Seymour (2005)  
³General references above on value of collaborative learning would also apply here, but in the context of teacher knowledge, skills, and metacognition.  
Sadler et al. (2013) |

¹Note that the item descriptions are abbreviated to save space. The full version of inventory in the Appendix should be consulted to fully understand what that item on the survey is asking. The classification is for the convenience of the reader rather than any sort of factor analysis. Many of the practices represented by a single inventory item contribute via several of the factors listed, and the factors themselves are not orthogonal. We list practices according to a somewhat arbitrary choice as to their single “most important” factor and the most relevant references, noting in italics some of the most important other factors by which that practice contributes. The references listed are not an exhaustive list and in most cases are reviews that contain many original references. This table does not include 14 commonly used teaching practices that are captured by the inventory to characterize the teaching methods used but are not given points in the scoring rubric due to insufficient evidence as to their impact on learning. Superscript numbers in column 2 refer to applicable references in column 3.

RESULTS FROM TYPICAL COURSES AND DEPARTMENTS

We present the results from the 179 completed version 2 inventories for courses in five science and mathematics departments during one semester.² It took instructors an average of 13 min (SD 6 min) to complete the inventory, with 53% of them taking 10 min or less. This is an overestimate of the time needed to fill out the inventory, as these times include the additional time to fill out three additional open-ended optional questions on institutional issues.³

²Although this is not identical to the version of the inventory shown in the Supplemental Material, we are confident these results would be nearly identical if that version had been used, as the changes from version 2 are very small, about half a dozen very minor word changes and the addition of one rarely chosen response option.

³The last question on the inventory asks how long it took to fill out the inventory. The time required also included the time faculty needed to respond to three additional UBC-specific open-ended questions:

- What do you see as the biggest barrier to achieving more effective student learning in the courses you teach? What changes could be made at UBC to help you teach more effectively? What changes could be made at UBC to increase your satisfaction/enjoyment of teaching?
- Approximately half the faculty members chose to provide answers to some or all of those questions.

These results are illustrative samples and do not represent an accurate characterization of all of these departments, because the level of response varied. Departments D2 and D3 made this a departmental priority and so obtained responses for 90% or more of their courses for the semester. The response rate of D1 is roughly 75%, D4 is well under 50%, and D5 is roughly 65%. Instructors who have worked with the CWSEI are disproportionately represented in these samples, and thus it is likely that the nonresponders from these departments would score lower than the departmental averages listed here.

In Table 3, we show the aggregate ETP scores and SDs for the five departments, including the breakdown of scores for...
each of the eight categories. Figure 1 shows the histograms of the ETP scores for the five departments. In Supplemental Table S1, we show the total and category ETP scores for each of the 31 courses in a single department as an example. The tables and figure provide an indication of the large amount of useful information provided by the inventory.

Figure 1 shows there is a substantial distribution of ETP scores within departments, covering a range of more than a factor of four in four of the departments and more than a factor of three in the fifth department. The lowest-scoring courses are at 10 and 11, while the highest-scoring courses are just above 50. This demonstrates that the inventory provides a large degree of discrimination between the practices of different instructors within a department. The spreads within departmental averages are larger than the differences between departmental averages. The category averages shown in Table 3 show larger fractional differences between departments than do the total ETP scores for departments.

In addition, we know that D1 has chosen to focus more than the other departments on introducing research-based practices into its largest courses. Consequently, as shown in Table 3, the difference between its average ETP and its enrollment-weighted average ETP is larger than the corresponding differences for the other departments.

The scored TPI also provides a measure of change in practices. We calculated the average ETP score and enrollment-weighted average score for department D3 for the 2006–2007 and 2012–2013 academic years using the 80% of the scored questions that were unchanged between the two versions of the inventory. Figure 2 and Table 4 show there has been a large change in these scores (about one SD, p < 0.001 using a simple t test) over this 6-yr period. It is notable that the TPI results show greater use of research-based practices in categories I, III, and IV in the later period (all differences statistically significant)—those are the categories in which the majority of CWSEI work in that department has been focused. Furthermore, the large fractional change in VIII is consistent with independent indications of increasing collaboration on teaching within the department. These results demonstrate that the inventory can capture both general and specific changes over time in the teaching practices within a department.

As a test of the TPI and scoring rubric, we sought out courses in which the instructors had published work showing they had spent multiple years investigating the use of different teaching practices while carefully monitoring student outcomes, and had achieved notable improvement. We found six such courses, spread across five institutions (two major research universities, two relatively small comprehensive universities, and one middle-tier research university) and three disciplines, including biology. We asked the six instructors to fill out the inventory. The hypothesis was that, if the ETP scores were capturing all or nearly all practices important for learning, these courses would have high ETP scores, providing another indication of validity, but if these courses had mid-range or lower ETP scores, it would indicate there must be important contributions to effective teaching that the TPI was missing. All of the instructors completed the inventory for us. Their ETP scores ranged from 46 to 57. The lowest of this set are extremely high compared with most courses for which we have data, while the top two scores are the highest we have ever seen. It is particularly notable that these top two scores, 54 and 57, were obtained in courses at comprehensive universities that have students with highly variable and relatively weak preparations and where the documented improvement in student outcomes achieved (Coletta, 2013; Adams et al., 2014) due to these multiyear efforts is spectacular.

**IMPLICATIONS FOR INCREASING THE USE OF RESEARCH-BASED PRACTICES**

Although the ETP score is useful for making general comparisons, the detailed information contained in the individual inventory responses by course is more useful for guiding improvements of teaching. The inventory can be valuable for instructors to use on their own for evaluating and improving their teaching; they can identify practices that they are not using and that have been shown to improve learning. For instructors who have made a serious effort to introduce research-based practices into their teaching, the inventory also provides a way they can quantify this for merit review or promotion.

At a departmental level, the TPI information reveals when there are instructors who are at odds with specific departmental or institutional norms. For example, it was a revelation to one department to learn that one long-time instructor did not employ graded homework assignments, while everyone else in the department did so automatically. As another example,
categories that are the norm for the department as a whole. For example, Table S1 shows that course 2 is relatively high in all categories except for the course information it provides, while course 9 is high in most areas but is unusually low in terms of in-class activities. It can also be seen that in categories I and VII most of the courses score fairly well, but a few are noticeably lower than the average. We examined the full spreadsheet showing individual item responses for all the courses (which is too massive to include with this paper) to understand more precisely how these courses were different. The differences came from the combination of the lack of learning objectives provided to students and the lack of regular coordination and guidance meetings with the TAs. These are two practices that are both desirable and widely embraced by the department. These examples illustrate how the information provided by this inventory reveals straightforward and efficient ways to improve the teaching of courses in a department.

FURTHER WORK

We suspect that the inventory will be valid for use in other disciplines, at least in the engineering and social sciences. This is based on our impression that the teaching practices used in these disciplines are rather similar to those used in math and science. It would be straightforward to check that the wording of the items would be correctly interpreted by instructors from those disciplines and that the inventory includes the teaching practices used in those disciplines. As needed, items could be reworded and added. There are reasonable justifications for most of the scoring rubric that transcend the specific disciplines of math and science (Bransford et al., 2000; Pascarella and Terenzini, 2005; Ambrose et al., 2010).

It would be valuable to go beyond simply capturing whether or not a practice is used and determine the quality of implementation of that practice. We have studied the difficulties in reliably measuring the quality of implementation of the practices being used and found them to be very formidable—far more difficult than determining what practices were being used. In 2011–2012, the CWSEI had a team of ~20 SESs who were experts (typically PhDs) in their respective science disciplines and who have had extensive training in science education research and practices (Wieman et al., 2010). They also had years of experience working with instructors and observing instructors teaching almost daily. They had interacted extensively with the students in the classes in which teaching practices were being changed, measuring their learning, interest, and engagement in a variety of ways. In short, they are far more qualified observers than anyone available to a typical academic department. These SESs were given the challenge of observing classes with which they were not familiar and evaluating the quality with which the instructor was implementing the teaching practices used. After trying to do this, the SESs concluded that they could not do so, except for detecting blatantly bad practices. Their conclusion was that to do a good evaluation of the quality with which the respective teaching practices are being used not only requires high levels of expertise in both the subject being taught and the teaching methods being used, but also considerable familiarity with the student population.
enrolled in the course. Fortunately, there is evidence showing that when regular instructors adopted research-based practices, the learning outcomes of their students substantially improved (Hake, 1998; Knight and Wood, 2005; Derting and Ebert-May, 2010; Hoellwarth and Moelter, 2011; Porter et al., 2013). We have also observed this many times in the CWSEI.

As much more extensive data are gathered on the teaching practices in use in STEM courses, for example, by widespread use of the TPI, it will be possible to carry out a more detailed analysis of the correlation between different practices and student outcomes under a range of conditions. This will allow a more refined scoring rubric to be created that is more precisely related to student outcomes. This will also allow the inventory to be refined to better capture what qualifies as effective use of a specific practice. For example, are there particular features that would make one student discussion board more beneficial than another, or are there certain midterm evaluation questions that evidence will show are particularly beneficial?

Another research direction that we are currently pursuing with collaborators is the development of a student version of the inventory. Comparisons of students’ and instructors’ inventory responses for the same course would likely provide valuable data on the students’ educational experiences and the level of communication between students and instructors.

SUMMARY

We have presented an inventory of teaching practices that provides a rich and detailed picture of what practices are used in a course. We also have presented a scoring rubric that gives a quantitative measure of the extent of use of research-based teaching practices that have been shown to result in improved student learning. We believe that this instrument will be a valuable tool for evaluating and improving undergraduate STEM teaching and, after further validation studies, will likely be useful in other disciplines as well. This inventory and scoring will need to be periodically updated to reflect future research on pedagogy and learning.

Table 4. Comparison of the teaching practices inventory data for the 2006–2007 and 2012–2013 academic years

<table>
<thead>
<tr>
<th></th>
<th>AVE (SD)</th>
<th>EWA</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3 2006–2007</td>
<td>20.4 (6.2)</td>
<td>19.2</td>
<td>2.3</td>
<td>3.4</td>
<td>2.9</td>
<td>2.5</td>
<td>6.0</td>
<td>0.7</td>
<td>0.8</td>
<td>2.0</td>
</tr>
<tr>
<td>D3 2012–2013</td>
<td>27.3 (6.8)</td>
<td>28.9</td>
<td>4.4</td>
<td>3.8</td>
<td>4.5</td>
<td>3.5</td>
<td>5.5</td>
<td>1.2</td>
<td>0.9</td>
<td>3.5</td>
</tr>
</tbody>
</table>

*Average and SD (AVE (SD)), enrollment-weighted average (EWA), and category averages for department D3. The scoring is lower than in Table 3, because it is based only on the subset of 40 scored questions common to both versions of the inventory. SEs for the category scores are 0.5 for category III and 0.3 for all the others.*
Appendix 1. Inventory showing formatting, with scoring and footnotes to references that justify the scoring. We did not insert the references directly in the document to allow the format to be shown. The formatting improves the user-friendliness of the inventory. A clean copy of the inventory is available at www.cwsei.ubc.ca/resources/TeachingPracticesInventory.htm.

Teaching Practices Inventory

(Scoring rubric points are the numbers in bold to right of each item.)

I. Course information provided to students via hard copy or course webpage. (check all that occurred in your course)  

☐ List of topics to be covered 1
☐ List of topic-specific competencies (skills, expertise, ...) students should achieve (what students should be able to do) 3
☐ List of competencies that are not topic related (critical thinking, problem solving, ...) 1
☐ Affective goals – changing students’ attitudes and beliefs (interest, motivation, relevance, beliefs about their competencies, how to master the material) 1
☐ Other (please specify)
   If you selected other, please specify

II. Supporting materials provided to students (check all that occurred in your course)  

☐ Student wikis or discussion boards with little or no contribution from you 0
☐ Student wikis or discussion boards with significant contribution from you or TA 1
☐ Solutions to homework assignments 1
☐ Worked examples (text, pencast, or other format) 1
☐ Practice or previous year’s exams 1
☐ Animations, video clips, or simulations related to course material 1
☐ Lecture notes or course PowerPoint presentations (partial/skeletal or complete) 1
☐ Other instructor selected notes or supporting materials, pencasts, etc. 0
☐ Articles from scientific literature 1
☐ Other (please specify)
   If you selected other, please specify

III. In-class features and activities  

A. Various  

Give approximate average number:

Average number of times per class: pause to ask for questions  (1 if >3)
Average number of times per class: have small group discussions or problem solving\(^1\)  

(1 if 1, 2 if >1)

Average number of times per class: show demonstrations, simulations, or video clips  

0

Average number of times per class: show demonstrations, simulations, or video where students first record predicted behavior and then afterwards explicitly compare observations with predictions\(^9\)  

(1 if >0.5)

Average number of discussions per term on why material useful and/or interesting from students' perspective\(^n\)  

(1 if 3-5, 2 if>5)

Comments on above (if any):__________

Check all that occurred in your course:

- Students asked to read/view material on upcoming class session  
  0

- Students read/view material on upcoming class session and complete assignments or quizzes on it shortly before class or at beginning of class\(^1\)  
  2

- Reflective activity at end of class, e.g. “one minute paper” or similar (students briefly answering questions, reflecting on lecture and/or their learning, etc.)\(^1\)  
  1

- Student presentations (oral or poster)\(^k\)  
  1

Fraction of typical class period you spend lecturing (presenting content, deriving mathematical results, presenting a problem solution, ...)\(^k\)  

(2 if 0-60%, 1 if 60-80%, 0 if 80-100%)

- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

Considering the time spent on the major topics, approximately what fraction was spent on the process by which the theory/model/concept was developed?\(^i\)  

(1 if more than 10%)

- 0-10%
- 10-25%
- more than 25%

---

\(^1\) Promising Practice No. 2: Organize Students in Small Groups in Froyd (2008); chap. 5 in Ambrose et al. (2010).

\(^2\) (Crouch et al., 2004; Sokoloff & Thornton, 1997, 2004).

\(^8\) Promising Practice No. 4: Scenario-based Content Organization in Froyd (2008); chap. 3 in Ambrose et al. (2010); (Pintrich, 2003).

\(^9\) (Novak et al., 1999); Although there is little peer-reviewed research showing the specific benefits of pre-class reading with associated quizzes, essentially every instructor that introduces active learning techniques in their classrooms reports that results are improved when they introduce pre-class reading. Similarly, when instructors in the Science Education Initiative give graded quizzes on the pre-class readings, we have always seen improvement in the fraction of students doing the reading.

\(^i\) (Froyd, 2008; Pascarella & Terenzini, 2005).

\(^n\) Promising Practice No. 6: Designing In-class Activities to Actively Engage Students in Froyd (2008); chap. 5 in Ambrose et al. (2010).

\(^k\) (Abd-El-Khalick & Lederman, 2000).
B. Personal Response System

If a student response system is used to collect responses from all students IN REAL TIME IN CLASS, what method is used? (check all that occurred in your course)
- electronic ("clickers") with student identifier 0
- electronic anonymous 0
- colored cards 0
- raising hands 0
- written student responses that are collected and reviewed in real time 0
- Other (please specify)

If you selected other, please specify

Number of questions posed followed by student-student discussion per class^ 2 if >1
Number of times used as quiz device (counts for marks and no student discussion) per class 0

IV. Assignments (check all that occurred in your course)
- Problem sets/homework assigned or suggested but did not contribute to course grade 0
- Problem sets/homework assigned and contributed to course grade at intervals of 2 weeks or less\(^\text{a}\) 2
- Paper or project (an assignment taking longer than two weeks and involving some degree of student control in choice of topic or design)\(^\text{b}\) 1
- Encouragement and facilitation for students to work collaboratively on their assignments\(^\text{b}\) 2
- Explicit group assignments\(^\text{b}\) 1
- Other (please specify)

If you selected other, please specify

V. Feedback and testing; including grading policies (check all that occurred in your course)

A. Feedback from students to instructor during the term\(^\text{c}\)
- Midterm course evaluation 1
- Repeated online or paper feedback or via some other collection means such as clickers 1
- Other (please specify)

If you selected other, please specify

B. Feedback to students (check all that occurred in your course)\(^\text{d}\)
- Assignments with feedback before grading or with opportunity to redo work to improve grade 2
- Students see marked assignments 1
- Students see assignment answer key and/or marking rubric 1
- Students see marked midterm exam(s) 1
- Students see midterm exam(s) answer key(s) 1
- Students explicitly encouraged to meet individually with you 1
- Other (please specify)

If you selected other, please specify

---
\(^a\) Promising Practice Nos. 2 & 6: Organize Students in Small Groups & Designing In-class Activities to Actively Engage Students in Froyd (2008); chap. 5 in Ambrose et al. (2010).
\(^b\) Chap. 5 in Ambrose et al. (2010); (Cooper et al., 2006; Walberg et al., 1985). The reviews by Cooper (2006) and Walberg (1985) are of the extensive K-12 research literature on the beneficial effects of graded homework. No such reviews exist for the study of homework in undergraduate math and science, but numerous articles report the educational benefit at this level. Two examples are Richards-Babb et al. (2011) and Cheng (2004).
\(^c\) (Kuh, 2008).
\(^d\) Promising Practice No. 2: Organize Students in Small Groups in Froyd (2008).
\(^e\) (Centra, 1973; Cohen, 1980; Diamond, 2004).
\(^f\) (Black & William, 1998; Hattie & Timperley, 2007); Promising Practice No. 5: Providing Students Feedback through Systematic Formative Assessment in Froyd (2008); chap. 5 in Ambrose et al. (2010).
C. Testing and grading

Number of midterm exams

Approximate fraction of exam mark from questions that required students to explain reasoning

Approximate breakdown of course mark (% in each of the following categories) 1 if final ≤60%, 0 if >60%

Final Exam
Midterm Exam(s)
Homework assignments
Paper(s) or project(s)
In-class activities
In-class quizzes
Online quizzes
Lab component
Participation
Other

If you selected other, please specify:

VI. Other (check all that occurred in your course)

☐ Assessment given at beginning of course to assess background knowledge† 1
☐ Use of instructor-independent pre-post test (e.g. concept inventory) to measure learning 2
☐ Use of a consistent measure of learning that is repeated in multiple offerings of the course to compare learning 2
☐ Use of pre-post survey of student interest and/or perceptions about the subject† 1
☐ Opportunities for students’ self-evaluation of learning ‡ 1
☐ Students provided with opportunities to have some control over their learning, such as choice of topics for course, paper, or project, choice of assessment methods, etc. ‡ 1
☐ New teaching methods or materials were tried along with measurements to determine their impact on student learning 2

VII. Training and guidance of Teaching Assistants (check all that occurred in your course) ‡

☐ No TAs for course 3 (to normalize)
☐ TAs must satisfy English language skills criteria † 1
☐ TAs receive ½ day or more of training in teaching 1
☐ There are Instructor–TA meetings every two weeks or more frequently where student learning and difficulties, and the teaching of upcoming material are discussed. 2
☐ TAs are undergraduates 0
☐ TAs are graduate students 0
☐ Other (please specify)

If you selected other, please specify

† (Gibbs & Simpson, 2005).
‡ Chap. 1 in Ambrose et al. (2010); chap. 1 in Bransford et al. (2000).
§ Chap. 7 in Ambrose et al. (2010); chap. 3 in Bransford et al. (2000).
° (Pintrich, 2003); chap. 3 in Ambrose et al. (2010).
‡ (Seymour, 2005).
§ (Anderson-Hsieh & Koehler, 1988; Hinnokota & Bailey, 1981; Jacobs & Friedman, 1988; Williams, 1992). There is little recent research on this, likely, because the use of language proficiency requirements for TAs has become so common.
VIII. Collaboration or sharing in teaching

☐ Used or adapted materials provided by colleague(s)   0
☐ Used "Departmental" course materials that all instructors of this course are expected to use7  1

Discuss how to teach the course with colleague(s)  1 if ≥3, 0 otherwise

☐ 1 Never
☐ 2
☐ 3
☐ 4
☐ 5 Very Frequently

Read literature about teaching and learning relevant to this course8  2 if ≥3, 1 if 2, 0 otherwise

☐ 1 Never
☐ 2
☐ 3
☐ 4
☐ 5 Very Frequently

Sat in on colleague’s class (any class) to get/share ideas for teaching 2 if ≥3, 1 if 2, 0 otherwise

☐ 1 Never
☐ 2
☐ 3
☐ 4
☐ 5 Very Frequently

IX. General (open-ended comments)

Please write any other comments here. If this inventory has not captured an important aspect of your teaching of this course, or you feel like you need to explain any of your above answers please describe it here.

Approximately how long did it take you to fill out this inventory?  

We thank you for taking the time to fill out this inventory.

7 There are many reports in the literature of collaborative efforts in teaching undergraduate science and mathematics. These are all local efforts, and the outcome measures are the self-reports of the participants. The many reports that we have examined all report perceived improvements in their teaching, but we know of no studies of the impact on student learning. Nevertheless, it is very likely that such collaborative efforts in teaching result in improved teaching, for much the same reason that collaborative activities with students result in improved learning, for which there is extensive evidence.

8 Common sense would suggest that such departmental materials are likely to receive much better vetting than those prepared by a teacher in isolation. This is certainly true for all of the numerous examples we know of across multiple institutions.

(Sadler et al., 2013). Although this reference is not on the direct value of reading the relevant science education literature, the article does show the benefit to student learning of instructor pedagogical content knowledge; knowledge that would plausibly be gained by reading the relevant literature.
ACKNOWLEDGMENTS

We are happy to acknowledge the assistance of all of the CWSEI SESS in this work, as well as the CWSEI departmental directors and many other UBC instructors who provided input. We are particularly grateful to Francis Jones and Brett Gilley for collecting the COPUS data. Michelle Smith, Peter Lepage, and Susan Singer provided helpful suggestions. This work was supported by UBC through the CWSEI.

REFERENCES


APPENDIX 1: APPROACHES TO TEACHING INVENTORY

This inventory is designed to explore the way that academics go about teaching in a specific context or subject or course. This may mean that your responses to these items in one context may be different to the responses you might make on your teaching in other contexts or subjects. For this reason we ask you to describe your context.

Please describe the subject/year of your response here: .........................................................

For each item please circle one of the numbers (1-5). The numbers stand for the following responses:
1 - this item was only rarely true for me in this subject.
2 - this item was sometimes true for me in this subject.
3 - this item was true for me about half the time in this subject.
4 - this item was frequently true for me in this subject.
5 - this item was almost always true for me in this subject.

Please answer each item. Do not spend a long time on each: your first reaction is probably the best one.

<table>
<thead>
<tr>
<th>Item</th>
<th>Statement</th>
<th>Only rarely</th>
<th>Almost always</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I design my teaching in this subject with the assumption that most of the students have very little useful knowledge of the topics to be covered.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I feel it is important that this subject should be completely described in terms of specific objectives relating to what students have to know for formal assessment items.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>In my interactions with students in this subject I try to develop a conversation with them about the topics we are studying.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>I feel it is important to present a lot of facts to students so that they know what they have to learn for this subject.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>I feel that the assessment in this subject should be an opportunity for students to reveal their changed conceptual understanding of the subject.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I set aside some teaching time so that the students can discuss, among themselves, the difficulties that they encounter studying this subject.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>In this subject I concentrate on covering the information that might be available from a good textbook.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>I encourage students to restructure their existing knowledge in terms of the new way of thinking about the subject that they will develop.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>In teaching sessions for this subject, I use difficult or undefined examples to provoke debate.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>I structure this subject to help students to pass the formal assessment items.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I think an important reason for running teaching sessions in this subject is to give students a good set of notes.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>In this subject, I only provide the students with the information they will need to pass the formal assessments.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>I feel that I should know the answers to any questions that students may put to me during this subject.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>I make available opportunities for students in this subject to discuss their changing understanding of the subject.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>I feel that it is better for students in this subject to generate their own notes rather than always copy mine.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>I feel a lot of teaching time in this subject should be used to question students' ideas.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

Thank you
Semi-Structured Interviews:

The following questions have been developed to use for semi-structured interviews with faculty. More items may be developed prior to the actual interview, but they will fall into the categories described.

Table 1. Sample Interview Items for Semi-Structured Interviews with Participating Faculty.

<table>
<thead>
<tr>
<th>Events</th>
<th>How do you intend to use/use the AACR assessments and reports?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patterns and Trends</td>
<td>Before you agreed to be part of the AACR group, had you ever heard of AACR assessments? In what context? How did this influence you? What appeals to you about the AACR assessments and reports? How will you know if you’re using the AACR assessments and reports effectively? What else do you need to know about the AACR assessments and reports in order to continue using them?</td>
</tr>
<tr>
<td>Systemic Structures</td>
<td>Outside of local hub meetings, have you interacted with members of the group about the AACR assessments or reports? If so, what did you talk about? Has this been useful? How do you expect your colleagues to react to your use of AACR assessments? How do you expect your administrators to react to your use of AACR assessments?</td>
</tr>
<tr>
<td>Mental Models</td>
<td>Have you seen any evidence regarding the utility of the AACR assessments? How do you know if you’re using AACR assessments and reports effectively? What value do you expect AACR assessments and reports to add to your teaching? How do the AACR assessments and reports impact your views of student learning?</td>
</tr>
</tbody>
</table>
**Interview and Survey Items for Administrators** - To be expanded and revised before data collection begins

Interview items will be open-ended. Survey items will be administered using a Likert scale.

What appeals to you about your faculty participating in the AACR project?

To what extent do you think people at your institution practice reformed teaching?

What do you see as the role of the instructor in the college science classroom?

What do you see as the role of the student in the college science classroom?

What do you see as the role of constructed-response assessments in the classroom?

In what ways does your institution provide administrative support for science instructors who want to try reformed teaching methods?

Have you seen any evidence regarding the utility of the AACR assessments?

How will you know if your faculty are implementing the AACR assessment effectively in their courses?

More generally, how do you evaluate faculty teaching effectiveness?
Survey of Undergraduate Teaching Practices

INFORMATION
This survey was designed by researchers at Western Michigan University to collect self-reported teaching practices from individuals teaching at institutions of higher education.

INSTRUCTIONS
The survey consists of twenty-four questions. It should take about 10 minutes to complete.

Each item is a statement that may represent your current teaching practice. As you proceed through the survey, please consider the statements as they apply to teaching your lowest level, largest enrollment undergraduate course taught in the last two years.

Please read each statement, then indicate the degree to which the statement is descriptive of your teaching. There are no “right” or “wrong” answers. The purpose of the survey is to understand how you teach, not to evaluate your teaching.

1 - Not at all descriptive of my teaching
2 - Minimally descriptive of my teaching
3 - Somewhat descriptive of my teaching
4 - Mostly descriptive of my teaching
5 - Very descriptive of my teaching
Questions about your course.
Please consider the lowest level, largest enrollment undergraduate course you are currently teaching or have taught in the last two years as you answer the following questions.

C1. Course Enrollment: _______ students

    ____% Majors in your discipline

    ____% Majors in other disciplines

C2. Is this a general education course? Yes / No / Not Applicable

C3. Weekly contact hours you teach per section:

   Lecture: ______________

   Lab: ______________

   Combined Lecture/Lab: ______________

   Discussion/Recitation: ______________

   Other (please specify): ______________

C4. If you think we need more information about your class, please explain:

C5. How are most decisions about teaching practices made?

    ____ I make most decisions

    ____ I’m part of a team that makes most decisions

    ____ Somebody else makes most decisions

Describe if applicable:
Optional Questions.
If you teach lecture and/or integrated lab, please indicate what proportion class time during a typical week is spent in the following activities. The sum of these answers should equal 100%.

P1. The instructor talking to the whole class. ______ %

P2. Students working individually. ______ %

P3. Students working in small groups. ______ %

P4. Students doing something else. (please specify) ______ % Other Activity: ____________  
    ______ % Other Activity: ____________  
    ______ % Other Activity: ____________
**Directions:** Please indicate the degree to which the following statements are descriptive of your teaching in your *lowest level, largest enrollment undergraduate course taught in the last 2 years*.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Not at all descriptive of my teaching</th>
<th>Minimally descriptive of my teaching</th>
<th>Somewhat descriptive of my teaching</th>
<th>Mostly descriptive of my teaching</th>
<th>Very descriptive of my teaching</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I guide students through major topics as they listen and take notes.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>2. I design activities that connect course content to my students’ lives and future work.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>3. My syllabus contains the specific topics that will be covered in every class session.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>4. I provide students with immediate feedback on their work during class (e.g., student response systems, short quizzes, etc.)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>5. I structure my course with the assumption that most of the students have little knowledge of the topics.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>6. I use student assessment results to guide the direction of my instruction during the semester.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>7. I frequently ask students to respond to questions during class time.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>8. I use student questions and comments to determine the focus and direction of classroom discussion.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>9. I have students use a variety of means (models, drawings, graphs, symbols, simulations, etc.) to represent phenomena.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
</tbody>
</table>
Directions: Please indicate the degree to which the following statements are descriptive of your teaching in your lowest level, largest enrollment undergraduate course taught in the last 2 years.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Not at all descriptive of my teaching</th>
<th>Minimally descriptive of my teaching</th>
<th>Somewhat descriptive of my teaching</th>
<th>Mostly descriptive of my teaching</th>
<th>Very descriptive of my teaching</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. I structure class so that students explore or discuss their understanding of new concepts before formal instruction.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>11. My class sessions are structured to give students a good set of notes.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>12. I structure class so that students regularly talk with one another about course concepts.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>13. I structure class so that students constructively criticize one another's ideas.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>14. I structure class so that students discuss the difficulties they have with this subject with other students.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>15. I require students to work together in small groups.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>16. I structure problems so that students consider multiple approaches to finding a solution.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>17. I provide time for students to reflect about the processes they use to solve problems.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>18. I give students frequent assignments worth a small portion of their grade.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
<tr>
<td>19. I require students to make connections between related ideas or concepts when completing assignments.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>D/K</td>
</tr>
</tbody>
</table>
### Directions: Please indicate the degree to which the following statements are descriptive of your teaching in your lowest level, largest enrollment undergraduate course taught in the last 2 years.

<table>
<thead>
<tr>
<th>Statement</th>
<th>1 Not at all descriptive of my teaching</th>
<th>2 Minimally descriptive of my teaching</th>
<th>3 Somewhat descriptive of my teaching</th>
<th>4 Mostly descriptive of my teaching</th>
<th>5 Very descriptive of my teaching</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. I provide feedback on student assignments without assigning a formal grade.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D/K</td>
</tr>
<tr>
<td>21. My test questions focus on important facts and definitions from the course.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D/K</td>
</tr>
<tr>
<td>22. My test questions require students to apply course concepts to unfamiliar situations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D/K</td>
</tr>
<tr>
<td>23. My test questions contain well-defined problems with one correct solution.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D/K</td>
</tr>
<tr>
<td>24. I adjust student scores (e.g., curve) when necessary to reflect a proper distribution of grades.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D/K</td>
</tr>
</tbody>
</table>
Instructors and the teaching practices they employ play a critical role in improving student learning in college science, technology, engineering, and mathematics (STEM) courses. Consequently, there is increasing interest in collecting information on the range and frequency of teaching practices at department-wide and institution-wide scales. To help facilitate this process, we present a new classroom observation protocol known as the Classroom Observation Protocol for Undergraduate STEM or COPUS. This protocol allows STEM faculty, after a short 1.5-hour training period, to reliably characterize how faculty and students are spending their time in the classroom. We present the protocol, discuss how it differs from existing classroom observation protocols, and describe the process by which it was developed and validated. We also discuss how the observation data can be used to guide individual and institutional change.

INTRODUCTION

A large and growing body of research indicates that undergraduate students learn more in courses that use active-engagement instructional approaches (Prince, 2004; Knight and Wood, 2005; Michael, 2006; Blanchard et al., 2010). As a result, the importance of teaching science, technology, engineering, and mathematics (STEM) courses more effectively has been stressed in numerous reports, including the President’s Council of Advisors on Science and Technology Engage to Excel report (2012), the National Science Foundation/American Association for the Advancement of Science Vision and Change report (AAAS, 2010), and the National Research Council Discipline-Based Education Research report (Singer et al., 2012). Given these compelling, evidence-based recommendations and the recognized need for measures of teaching effectiveness beyond student evaluations (Association of American Universities, 2011), higher education institutions are struggling to determine the extent to which faculty members are teaching in an interactive manner. This lack of information is a major barrier to transforming instruction and evaluating the success of programs that support such change.

To collect information about the nature of STEM teaching practices as a means to support institutional change, faculty at both the University of British Columbia (UBC) and the University of Maine (UMaine) created classroom observation programs. The results of such observations were needed to: 1) characterize the general state of STEM classroom teaching at both institutions, 2) provide feedback to instructors who desired information about how they and their students were spending time in class, 3) identify faculty professional development needs, and 4) check the accuracy of the faculty reporting on the Teaching Practices Survey that is now in use at UBC (CWSEI Teaching Practices Survey, 2013).

To achieve these goals, the programs needed an observation protocol that could be used by faculty member observers.
to reliably characterize how students and instructors were spending their time in undergraduate STEM classrooms. A critical requirement of the protocol was that observers who were typical STEM faculty members could achieve those results with only 1 or 2 hours of training, as it is unrealistic to expect they would have more time than that available. In the quest for a suitable observation protocol, multiple existing options were considered, and ultimately rejected.

The observation protocols considered were divided into two categories: open-ended or structured. When observers use open-ended protocols, they typically attend class, make notes, and respond to such statements as: “Comment on student involvement and interaction with the instructor” (Millis, 1992). Although responses to these types of questions can provide useful feedback to observers and instructors, the data are observer dependent and cannot easily be standardized or compared across multiple classrooms (e.g., all STEM courses at UBC or UMaine).

Alternatively, structured protocols provide a common set of statements or codes to which the observers respond. Often, these protocols ask observers to make judgments about how well the teaching conforms to a specific standard. Examples of such protocols include the Inside the Classroom: Observation and Analytic Protocol (Weiss et al., 2003) and the Reformed Teaching Observation Protocol (RTOP; Sawada et al., 2002). These protocols consist of statements that observers typically score on a Likert scale from “not at all” to “to a great extent” and contain such statements as: “The teacher had a solid grasp of the subject matter content inherent in the lesson” (from RTOP; Sawada et al., 2002).

The RTOP in particular has been used to observe university STEM instruction. For example, it has been used to evaluate university-level courses at several different institutions to measure the effectiveness of faculty professional development workshops (Ebert-May et al., 2011) and to compare physics instructors in a study examining coteaching as a method to help new faculty develop learner-centered teaching practices (Henderson et al., 2011). The RTOP is also being used to characterize classroom practices in many institutions and in all levels of geoscience classes (Classroom Observation Project, 2011).

The RTOP was found to be unsuitable for the UBC and UMaine programs for two main reasons. The first is that the protocol involves many observational judgments that can be awkward to share with the instructor and/or the larger university community. The second is that observers must complete a multiday training program to achieve acceptable interrater reliability (IRR; Sawada et al., 2002).

More recently, new observation protocols have been developed that describe instructional practices without any judgment as to whether or not the practices are effective or aligned with specific pedagogic strategies. These observation protocols use a series of codes to characterize instructor and/or student behaviors in the classroom; observers indicate how often each behavior occurs during a class period (Hora et al., 2013; West et al., 2013). One observation protocol in particular, the Teaching Dimensions Observation Protocol (TDOP), was expressly developed to observe postsecondary nonlaboratory courses. For this protocol, observers document classroom behaviors in 2-min intervals throughout the duration of the class session (Hora et al., 2013). The possible classroom behaviors are described in 46 codes in six categories, and observers make a checkmark when any of the behaviors occur.

The TDOP instrument avoids the judgment issues associated with the RTOP, but it still requires substantial training, as one might expect for a protocol that was designed to be a complex research instrument. Preliminary work suggests that, after a 3-day training session, observers have acceptable IRR scores when using the TDOP (Hora et al., 2013). Observers at our institutions tried using this instrument, but without the full training, they found it difficult to use the TDOP in a reliable way, due to the complexity of the items being coded and the large number of possible behavior codes. We also found that the particular research questions it was designed to address did not entirely align with our needs. For example, it covers some aspects that are not necessary for faculty observation programs, such as whether an instructor uses instructional artifacts (e.g., a laser pointer or computer; Hora et al., 2013) and fails to capture others that are needed, such as whether an instructor encourages peer discussion along with clicker questions (Mazur, 1997; Smith et al., 2009, 2011). We also wanted to better characterize the student behaviors during the class period than the TDOP easily allowed.

Out of necessity, we created a new protocol called the Classroom Observation Protocol for Undergraduate STEM, or COPUS. Like the TDOP, this new protocol documents classroom behaviors in 2-min intervals throughout the duration of the class session, does not require observers to make judgments of teaching quality, and produces clear graphical results. However, COPUS is different in that it is limited to 25 codes in only two categories (“What the students are doing” and “What the instructor is doing”) and can be reliably used by university faculty with only 1.5 hours of training (Figure 1 has a description of the codes; the Supplemental Material includes the full protocol and coding sheet). Observers who range from STEM faculty members without a background in science education research to K–12 STEM teachers have reliably used this protocol to document instruction in undergraduate science, math, and engineering classrooms. Taken together, their results show the broad usability of COPUS.

**DEVELOPMENT**

The development of COPUS was an evolutionary process extending across more than 2 years, involving many iterations and extensive testing. It began at UBC, where science education specialists (SESs) who were working with science faculty on improving teaching (Wieman et al., 2010) wanted to characterize what both the students and instructors were doing during class. The SESs began testing various existing protocols, including the TDOP, in different classes at UBC in late 2011 and early 2012. The original TDOP did not meet our needs (as described above), so we iteratively modified the protocol through nine different versions. These changes resulted in a format, procedure, data structure, and coding strategy that was easy to implement on paper or electronically and convenient for analysis and display. The overall format of the observation protocol remained largely stable, but the categories and codes continued to evolve.

During the Fall term of 2012, 16 SESs, who are highly trained and experienced classroom observers, used this evolving protocol to observe a variety of courses in singles,
Observation Protocol for STEM Classes

1. Students are Doing

- **L** Listening to instructor/taking notes, etc.
- **Ind** Individual thinking/problem solving. Only mark when an instructor explicitly asks students to think about a clicker question or another question/problem on their own.
- **CG** Discuss clicker question in groups of 2 or more students
- **WG** Working in groups on worksheet activity
- **OG** Other assigned group activity, such as responding to instructor question
- **AnQ** Student answering a question posed by the instructor with rest of class listening
- **SQ** Student asks question
- **WC** Engaged in whole class discussion by offering explanations, opinion, judgment, etc. to whole class, often facilitated by instructor
- **Prd** Making a prediction about the outcome of demo or experiment
- **SP** Presentation by student(s)
- **TQ** Test or quiz
- **W** Waiting (instructor late, working on fixing AV problems, instructor otherwise occupied, etc.)
- **O** Other – explain in comments

2. Instructor is Doing

- **Lec** Lecturing (presenting content, deriving mathematical results, presenting a problem solution, etc.)
- **RtW** Real-time writing on board, doc. projector, etc. (often checked off along with Lec)
- **FUp** Follow-up/feedback on clicker question or activity to entire class
- **PQ** Posing non-clicker question to students (non-rhetorical)
- **CQ** Asking a clicker question (mark the entire time the instructor is using a clicker question, not just when first asked)
- **AnQ** Listening to and answering student questions with entire class listening
- **MG** Moving through class, guiding ongoing student work during active learning task
- **1to1** One-on-one extended discussion with one or a few individuals, not paying attention to the rest of the class (can be along with MG or AnQ)
- **DV** Showing or conducting a demo, experiment, simulation, video, or animation
- **Adm** Administration (assign homework, return tests, etc.)
- **W** Waiting when there is an opportunity for an instructor to be interacting with or observing/listening to student or group activities and the instructor is not doing so
- **O** Other – explain in comments

The protocol evolved through five different versions during this stage of testing and feedback. The final version had substantially simplified categories and all identified problems with the wording on the codes had been eliminated. Notably, it was quite simple to reliably code classes taught with traditional lectures, as a very small number of behaviors need to be coded. Therefore, the majority of the work went into improving the protocol so it could reliably characterize classes that had substantial and varied interactions between instructor and students and multiple student activities.

One substantial change during Fall 2012 was eliminating a category for judging the cognitive level of the activities. Observers had been asked to code the level of cognitive sophistication of current classroom activities, based on Bloom’s taxonomy of educational objectives (Bloom et al., 1956). After multiple unsuccessful attempts to find a simple and reliable coding scheme that could capture this aspect of the classroom activities, we dropped this category. Our decision to drop this category is supported by recent work showing that, when faculty members write and evaluate higher-order questions, they use several criteria beyond the Bloom’s level, including: question difficulty, time required to answer the questions, whether students are using a new or well-practiced approach, and whether the questions have multiple reasonable solutions (Lemons and Lemons, 2012).

The second substantial change during this time was changing another category—coding the level of student engagement—from required to optional. Having a measure of student engagement is useful for providing feedback to the instructor and for judging the overall effectiveness of many instructional activities. With the coding of the levels of engagement simplified to only discriminating between low (0–20% of the students engaged), medium, or high (≥80% of the student engaged), some observers, particularly those who had some experience with observing levels of student engagement, could easily code engagement along with the other two categories and there was reasonable consistency between observers. However, less-experienced observers found it quite hard to simultaneously code what the students were doing, what the instructor was doing, and the student engagement level. Also, there were difficulties with obtaining consistent coding of student engagement across all observers; the judgments were often dependent on the levels of engagement common to the specific disciplines and courses with which the observers were familiar. For this reason, the student engagement category was made optional. We recommend observers do not try to code it until after they have become experienced at coding the “What the students are doing” and “What the instructor is doing” categories.
Another recurring theme of the discussions with the SESs was the extent to which classroom observations could accurately capture the quality of instruction or the efficacy of student work. In the end, after SESs observed different classes across many disciplines, there was a consensus that accurately evaluating the quality of instruction and the efficacy of student work was generally not possible. These highly trained and experienced observers concluded that these evaluations require a high degree of training of the observer in the material and the pedagogic strategies, as well as familiarity with the student population (prior knowledge, typical classroom behaviors, etc.). We concluded that quality judgments of this type were not realistic goals for limited classroom observations carried out by STEM faculty members. Thus, the present version of COPUS captures the actions of both instructors and students, but does not attempt to judge the quality of those actions for enhancing learning.

After the completion of this development work at UBC, the COPUS was further tested by 16 K–12 teachers participating in a teacher professional development program at UMaine. The teachers used the COPUS to observe 16 undergraduate STEM courses in five different departments (biology, engineering, math, chemistry, and physics). While the teachers easily interpreted many of the codes, they found a few to be difficult and suggested additional changes. For example, the student code “Listening: paying attention/taking notes, etc.” was changed to “Listening to instructor/taking notes, etc.” The code was clarified, so observers knew they should select this code only when the students were listening to their instructor, not when students were listening to their peers. Also, new codes were added to capture behaviors the teachers thought were missing, such as the instructor code “AnQ: Listening to and answering student questions with entire class listening.”

The coding patterns of the two teacher observers in the same classroom were also compared to determine which specific codes were difficult to use consistently. An example comparing two teachers employing the student code “Ind” is shown in Figure 2. Figure 2A compares how two observers marked this code in the first iteration of testing, when it was described “Ind: Individual thinking/problem solving in response to assigned task.” Observer 2 marked this code throughout most of the class, and observer 1 marked this code intermittently. Follow-up conversations with observer 2 and other teachers indicated that some observers were marking this code throughout the duration of the class, because they assumed individual students were thinking while they were taking notes, working on questions, and so on, but other observers were not. Therefore, we clarified the code to be: “Ind: Individual thinking/problem solving. Only mark when an instructor explicitly asks students to think about a clicker question or another question/problem on their own.”
Table 1. Information on the courses observed using the final version of the COPUS

<table>
<thead>
<tr>
<th>Institution</th>
<th>Number of classes observed</th>
<th>Number of different STEM departments</th>
<th>Percentage of courses at the introductory level</th>
<th>Percentage of classes with &gt;100 students</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBC</td>
<td>8</td>
<td>4(^b)</td>
<td>100</td>
<td>63</td>
</tr>
<tr>
<td>UMaine</td>
<td>23</td>
<td>7(^c)</td>
<td>96</td>
<td>35</td>
</tr>
</tbody>
</table>

\(^a\) STEM courses at the first- and second-year levels.
\(^b\) Biology, chemistry, math, and physics.
\(^c\) Biology, molecular biology, engineering, chemistry, math, physics, and geology.

Figure 2B shows a comparison of the same observer pair, with the revised “Ind” code showing how the paired codes were now closely aligned.

In addition, the teacher observation data revealed a more general problem: there was a lower degree of consistency in coding student behaviors than in coding instructor behaviors, and the teachers used a very limited set of codes for the student behaviors. The earlier coding by the SESs had shown similar, but less dramatic, trends. We realized that this problem was due to a natural tendency of observers to focus on the instructor, combined with the fact the instructor-related codes came first on the survey form. Therefore, the protocol was changed, with the student codes viewed first, and we emphasized coding student behaviors during subsequent training sessions (see further details below in the Training section). As shown below, these changes appear to have fixed this problem.

These further revisions culminated in a final version of the COPUS. This version was tested by having the same 16 K–12 teachers use it to observe 23 UMaine STEM classes, and by having seven STEM faculty observer volunteers to observe eight UBC classrooms in pairs after 1.5 hours of training. Information about the types of classes observed is in Table 1. The seven UBC STEM faculty member volunteers who used the final protocol had not previously used the protocol and were not involved in the development process. Thus, the IRR of the protocol has been tested with a sample of observers with a wide range of backgrounds and perspectives. As discussed in Validity and Reliability, the IRR was high.

TRAINING

A critical design feature of the COPUS is that college and university faculty who have little or no observation protocol experience and minimal time for training can use it reliably.

We summarize the training steps in the following paragraphs, and we have also included a step-by-step facilitator guide in the Supplemental Material.

The first step in the training process is to have the observers become familiar with the codes. At UBC, facilitators displayed the student and instructor codes (Figure 1) and discussed with the observers what each behavior typically looks like in the classroom. At UMaine, the teacher observers played charades. Each teacher randomly selected a code description from a hat and silently acted out the behavior. The remaining observers had the code descriptions in front of them and guessed the code. The remainder of the training was the same for both groups, with a total training duration of 2 hours for the K–12 teachers and 1.5 hours for the UBC faculty members.

Second, observers were given paper versions of the coding sheet and practiced coding a 2-min segment of a classroom video. An excerpt from the coding sheet is shown in Figure 3, and the complete coding sheet is included in the Supplemental Material. Observers often mark more than one code within a single 2-min interval. The first video we used showed an instructor making administrative announcements and lecturing while the class listened. After 2 min, the video was paused, and the group discussed which codes they selected. Because faculty at other institutions may have difficulty capturing videos for training, we have included web URLs to various video resources that can be used for training (Table 2).

The observers were then asked to form pairs and code 8 min of a video from a large-enrollment, lecture-style science class at UMaine that primarily shows an instructor lecturing and students listening, with a few questions asked by both the instructor and students. To keep the observers synchronized and ensure they were filling out a new row in the observation protocol at identical 2-min intervals, they used either cell phones set to count time up or a sand timer. At

![Table 1](image)

**Figure 3.** An excerpt of the COPUS coding form. Observers place a single checkmark in the box if a behavior occurs during a 2-min segment. Multiple codes can be marked in the same 2-min block.
the end of 8 min, the observers compared their codes with their partners. Next, as a large group, observers took turns stating what they coded for the students and the instructor every 2 min for the 8-min video clip. At this point, the observers talked about the relationship between a subset of the student and instructor codes. For example, if the observers check the student code “CG: Discuss clicker question,” they will also likely check the instructor code “CQ: Asking a clicker question.”

To provide the observers with practice coding a segment that has more complicated student and instructor codes, they next coded a different classroom video segment from the same large-enrollment, lecture-style science class at UMaine, but this time the camera was focused on the students. This video segment included students asking the instructor questions, students answering questions from the instructor, and clicker questions with both individual thought and peer discussion. The observers coded 2 min and then paused to discuss the codes. Then observers in pairs coded for an additional 6 min, again taking care to use synchronized 2-min increments. The observer pairs first compared their codes with their partners, and then the whole group discussed the student and instructor codes for each of the 2-min segments of the 6-min clip. At this point, the training was complete.

### VALIDITY AND RELIABILITY

COPUS is intended to describe the instructor and student actions in the classroom, but it is not intended to be linked to any external criteria. Hence, the primary criterion for validity is that experts and observers with the intended background (STEM faculty and teachers) see it as describing the full range of normal classroom activities of students and instructors. That validity was established during the development process by the feedback from the SESs, the K–12 teachers, and those authors (M.S., F.J., C.W.) who have extensive experience with STEM instruction and classroom observations.

A major concern has been to ensure that there is a high level of IRR when COPUS is used after the brief period of training described above. To assess the IRR, we examined the agreement between pairs of observers as they used the final version of COPUS in STEM classes at both UBC and UMaine. The two observers sat next to each other in the classroom, so they could keep identical 2-min time increments, but the observers were instructed not to compare codes with each other.

To summarize how similarly observer pairs used each code on the final version of the COPUS, we calculated Jaccard similarity scores (Jaccard, 1901) for each code and then averaged the scores for both the UBC and UMaine observers (Table 3).

### Table 2. Video resources that may be helpful for COPUS training

<table>
<thead>
<tr>
<th>Description of video</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstration, clicker questions, and lecture</td>
<td><a href="http://harvardmagazine.com/2012/02/interactive-teaching">http://harvardmagazine.com/2012/02/interactive-teaching</a></td>
</tr>
<tr>
<td>Group activities and lecture</td>
<td><a href="http://ocw.mit.edu/courses/biology/7-012-introduction-to-biology-fall-2004/video-lectures/lecture-6-genetics-1">http://ocw.mit.edu/courses/biology/7-012-introduction-to-biology-fall-2004/video-lectures/lecture-6-genetics-1</a></td>
</tr>
<tr>
<td>Clicker, real-time writing, and lecture</td>
<td><a href="http://podcasting.gcsu.edu/4DCGI/Podcasting/UGA/Episodes/22253/27757327.mov">http://podcasting.gcsu.edu/4DCGI/Podcasting/UGA/Episodes/22253/27757327.mov</a></td>
</tr>
<tr>
<td>Real-time writing, asking/answering questions, and lecture</td>
<td><a href="http://podcasting.gcsu.edu/4DCGI/Podcasting/UGA/Episodes/12746/614158822.mov">http://podcasting.gcsu.edu/4DCGI/Podcasting/UGA/Episodes/12746/614158822.mov</a></td>
</tr>
</tbody>
</table>

### Table 3. Average Jaccard similarity scores for COPUS codes across all pairs observing in all courses for both UBC faculty observers and Maine K–12 teacher observers; numbers closer to 1 indicate the greatest similarity between two observers

<table>
<thead>
<tr>
<th>Student code</th>
<th>UBC</th>
<th>UMaine</th>
<th>Instructor code</th>
<th>UBC</th>
<th>UMaine</th>
</tr>
</thead>
<tbody>
<tr>
<td>L: Listening</td>
<td>0.95</td>
<td>0.96</td>
<td>Lec: Lecturing</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>Ind: Individual thinking/problem solving</td>
<td>0.97</td>
<td>0.91</td>
<td>RW: Real-time writing</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>CG: Discuss clicker question</td>
<td>0.98</td>
<td>0.97</td>
<td>FQ: Follow-up on clicker questions or activity</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td>WG: Working in groups on worksheet activity</td>
<td>0.98</td>
<td>0.99</td>
<td>PQ: Posing nonclicker questions</td>
<td>0.86</td>
<td>0.80</td>
</tr>
<tr>
<td>OG: Other group activity</td>
<td>Not used</td>
<td>0.97</td>
<td>CQ: Asking a clicker question</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td>AnQ: Students answer question posed by instructor</td>
<td>0.91</td>
<td>0.84</td>
<td>AnQ: Answering student questions</td>
<td>0.94</td>
<td>0.89</td>
</tr>
<tr>
<td>SQ: Student asks question</td>
<td>0.96</td>
<td>0.93</td>
<td>MG: Moving through the class</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>WC: Engaged in whole-class discussion</td>
<td>0.96</td>
<td>0.98</td>
<td>Io1: One-on-one discussions with students</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Prd: Making a prediction about the outcome of demo or experiment</td>
<td>Not used</td>
<td>1.00</td>
<td>D/V: Conducting a demo, experiment, etc.</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>SP: Presentation by students^a</td>
<td>Not used</td>
<td>Not used</td>
<td>Adm: Administration</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>TQ: Test or quiz^a</td>
<td>Not used</td>
<td>Not used</td>
<td>W: Waiting</td>
<td>0.95</td>
<td>0.98</td>
</tr>
<tr>
<td>W: Waiting</td>
<td>0.99</td>
<td>0.98</td>
<td>O: Other</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td>O: Other</td>
<td>0.94</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a“SP: Presentation by students” and “TQ: Test/quiz” were not selected in any of the observations at UBC or UMaine. This result likely occurred because when we asked UBC and UMaine faculty members if we could observe their classes, we also asked them if there was anything unusual going on in their classes that day. We avoided classes with student presentations and tests/quizzes, because these situations would limit the diversity of codes that could be selected by the observers.
For single codes, we calculated Jaccard similarity scores instead of IRR Cohen’s kappa values, because observer pairs occasionally marked the same code for every 2-min increment throughout the duration of the class. For example, in a class that is lecture-based, observers would likely mark the student code “L: Listening” for the entire time. In a case such as this, the observer opinion is defined as a constant rather than a variable, which interferes with the IRR calculation.

The equation for the Jaccard coefficient is \( T = \frac{n_a}{n_a + n_b - n_c} \), where \( n_a \) is the number of 2-min increments that are marked the same (either checked or not checked) for both observers, \( n_c \) is the number of 2-min increments that are marked the same for both observers plus 2-min increments observer 1 marked that observer 2 did not, \( n_b \) is number of 2-min increments that are marked the same for both observers plus 2-min increments observer 2 marked that observer 1 did not. For example, for the data in Figure 2B, the class period is 42 min in length, so there are 21 possible 2-min segments. The student code “Ind: Individual thinking” was marked 12 times by observers 1 and 2, not marked eight times by both observers, and marked by observer 2 one time when observer 1 did not. Therefore, the calculation is: \( 20 / (20 + 21 - 20) = 0.95 \). Numbers closer to 1 indicate greater consistency between how the two observers coded the class.

Eighty-nine percent of the similarity scores are greater than 0.90, and the lowest is 0.80. These values indicate strong similarity between how two observers use each code. The lowest score for both the UBC and UMaine observers was for the instructor code “PQ: Posing nonclicker questions.” Comments from observers suggest that, when instructors were following up/giving feedback on clicker questions or activities, they often posed questions to the students. Observers checked the instructor code “FU: Follow-up” to describe this behavior but stated they occasionally forgot to also select the instructor code “PQ.”

To compare observer reliability across all 25 codes in the COPUS protocol, we calculated Cohen’s kappa IRR scores using SPSS (IBM, Armonk, NY). To compute the kappa values for each observer pair, we added up the total number of times: 1) both observers put a check in the same box, 2) neither observer put a check in the same box, 3) observer 1 put a check in a box when observer 2 did not, and 4) observer 2 put a check in a box when observer 1 did not. For example, at UBC, when looking at all 25 codes in the COPUS, one observer pair had the following results: 1) both observers put a check in 83 of the same boxes, 2) neither observer put a check in 524 of the boxes, 3) observer 1 marked six boxes when observer 2 did not, and 4) observer 2 marked 12 boxes that observer 1 did not. Using data such as these, we computed the kappa score for each of the eight UBC and 23 UMaine pairs and report the average scores in Table 4. We also repeated this calculation using either the subset of 13 student or 12 instructor codes (Table 4).

The average kappa scores ranged from 0.79 to 0.87 (Table 4). These are considered to be very high values for kappa and thus indicate good IRR (Landis and Koch, 1977). Notably, the kappa values, as well as the Jaccard similarity scores, are comparably high for both UBC faculty and UMaine K–12 teacher observers, indicating that COPUS is reliable when used by observers with a range of backgrounds and 2 hours or fewer of training.

### ANALYZING COPUS DATA

To determine the prevalence of different codes in various classrooms, we added up how often each code was marked by both observers and then divided by the total number of codes shared by both observers. For example, if both observers marked “Instructor: Lecture” at the same 13 time intervals in a 50-min class period and agreed on marking 25 instructor codes total for the duration of the class, then 13/25, or 52% of the time, the lecture code occurred for the instructor.

We visualized the prevalence of the student and instructor codes using pie charts. Figure 4 shows observation results from two illustrative classes: one that is primarily lecture-based and one in which a combination of active-learning strategies are used. The latter class is clearly differentiated from the lecture-based class. This example illustrates how, at a glance, this visual representation of the COPUS results provides a highly informative characterization of the student and instructor activities in a class.

At a department- or institution-wide level, there are several ways to categorize the range of instructional styles. One of the simplest is to look at the prevalence of the student code “L: Listening to instructor/taking notes, etc.” across all courses observed, because this student code is the most indicative of student passive behavior in response to faculty lecturing (“Lec”) with or without real-time writing (“RtW”). Figure 5 shows that at both institutions the “L” code was marked 26–75% of the time. However, at UMaine, some of the classes have greater than 76% of the student codes devoted to listening. Faculty who teach these classes may benefit from professional development activities about how to design an effective active-learning classroom.

In addition, the data can be analyzed for a subset of faculty members who are using active-learning strategies, such as asking clicker questions. Thirty-eight percent of UBC and 43% of the UMaine classes that were observed used clickers. However, student code prevalence in these classes show that not all faculty members used clicker questions accompanied by recommended strategies, such as peer discussion (Mazur, 1997; Smith et al., 2009, 2011; Figure 6). Faculty members who are not allowing time for peer discussion may benefit from professional development on how to integrate peer discussion into clicker questions.

### Table 4. Average IRR kappa scores from the observations at UBC and UMaine

<table>
<thead>
<tr>
<th>Observers</th>
<th>All codes (± SE)</th>
<th>Student codes (± SE)</th>
<th>Instructor codes (± SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty observing UBC courses</td>
<td>0.83 (0.03)</td>
<td>0.87 (0.04)</td>
<td>0.79 (0.04)</td>
</tr>
<tr>
<td>Teachers observing UMaine courses</td>
<td>0.84 (0.03)</td>
<td>0.87 (0.04)</td>
<td>0.82 (0.04)</td>
</tr>
</tbody>
</table>

Vol. 12, Winter 2013
Figure 4. A comparison of COPSUS results from two courses that have different instructional approaches.

Figure 5. Prevalence of the student code “L: Listening” across several UBC and UMaine classes.

Figure 6. Prevalence of student codes in four example courses that use clickers. In courses that use clickers with no or minimal peer discussion, the students are passively listening the majority of the time.
DISCUSSION AND SUMMARY

COPUS was developed because university observation programs needed a protocol to: 1) characterize the general state of teaching, 2) provide feedback to instructors who desired information about how they and their students were spending class time, and 3) identify faculty professional development needs. COPUS meets all of these goals by allowing observers with little observation protocol training and experience to reliably characterize what both faculty and students are doing in a classroom.

There are several uses for COPUS data. On an individual level, faculty members can receive pie charts with their code prevalence results (examples in Figure 4). These results provide a nonthreatening way to help faculty members evaluate how they are spending their time. We discovered that faculty members often did not have a good sense of how much time they spent on different activities during class, and found COPUS data helpful.

In addition, faculty members can use COPUS data in their tenure and promotion documents to supplement their normal documentation, which typically includes student evaluation information and a written description of classroom practices. Having observation data gives faculty members substantially more information to report about their use of active-learning strategies than is usually the case.

COPUS data can also be used to develop targeted professional development. For example, anonymized, aggregate COPUS data across all departments have been shared with the UMaine Center for Excellence in Teaching and Assessment, so workshops and extended mentoring opportunities can better target the needs of the faculty. One area in particular that will be addressed in an upcoming professional development workshop is using clickers in a way that promotes peer discussion. The idea for this workshop came about as a result of the COPUS evidence showing the prevalence of UMaine STEM classes that were using clickers but allowing no or minimal time for recommended student peer discussions (Figure 6).

Other planned uses for COPUS include carrying out systematic observations of all instructors in a department at UBC in order to characterize teaching practices. The information will be used with other measures to characterize current usage of research-based instructional practices across the department’s courses and curriculum.

In the end, the choice of observation protocol and strategy will depend on the needs of each unique situation. COPUS is easy to learn, characterizes nonjudgmentally what instructors and students are doing during a class, and provides data that can be useful for a wide range of applications, from improving an instructor’s teaching or a course to comparing practices longitudinally or across courses, departments, and institutions.

ACKNOWLEDGMENTS

This work was supported at UBC through the Carl Wieman Science Education Initiative and by the National Science Foundation under grant #0962805. We are grateful for the assistance of all of the UBC SEs who contributed to the development of the survey; Lisa McDonnell and Bridgette Clarkson for running the UBC training session; MacKenzie Stetzer, Susan McKay, Erika Allison, Medea Steinman, and Joanna Meyer for helping to run the UMaine training session; Jeremy Smith for developing scripts to parse and analyze the data; the Maine K–12 teachers and UBC faculty volunteers who served as observers; and the faculty at UBC and UMaine who allowed their courses to be observed.

Approval to observe classrooms and instruction at UBC and publish results of that work is provided to the Carl Wieman Science Education Initiative by the University of British Columbia under the policy on institutional research. Approval to evaluate teacher observations of classrooms (exempt status, protocol no. 2013-02-06) was granted by the Institutional Review Board at the University of Maine.

REFERENCES


To help institutions collect information on undergraduate teaching practices, the authors developed a new classroom observation protocol known as the Classroom Observation Protocol for Undergraduate STEM (COPUS). This protocol allows college science, technology, engineering, and mathematics faculty, after a short training period, to reliably characterize how faculty and students are spending their time in class.