Supplemental Materials for Instructional Utility and Learning Efficacy of Common Active Learning Strategies

David A. McConnell[1], LeeAnna Chapman[1], C. Douglas Czajka[1], Jason P. Jones[1], Katherine D. Ryker[2], and Jennifer Wiggen[1]
[1] Department of Marine, Earth and Atmospheric Sciences, North Carolina State University
[2] Department of Geography and Geology, Eastern Michigan University

Development of Rubrics

In creating the rubric to evaluate each teaching strategy's utility (Table 2), it was not our intention to create a tool for others to use in evaluating teaching strategies. The rubric was based on our experience as a group of both geoscience education researchers and instructors, and was created to aid our evaluation of the strategies and frame our discussion of their utility or ease of use. The scoring rubric allowed us to rank the strategies in terms of utility based on a variety of factors, and we felt it would serve as a useful visual reference for readers to supplement the text discussion. The categories and scoring criteria were developed out of our experience conducting researching utilizing some of these strategies and using them in a variety of introductory geoscience courses (e.g., physical geology, earth science, environmental geology, historical geology) and class sizes (30-200+ students). Other instructors may come up with a different assessment of utility for the same strategies using this rubric based on their own experiences and situational factors. However, we feel that that iterative process of collaborative revision described below lends some evidence for content validity the rubric and that it would provide at least moderate reliability if used by others.

The initial list of scoring criteria for the rubric came out of a meeting with all the authors. Based on this draft list, each author was then assigned two strategies to independently evaluate with the criteria. No scoring guidelines were established for each criterion at first, so each author was descriptive in their assessment of each criterion. For example, instead of having three

defined scoring levels to rate required resources, a short description of the required resources was written instead. After this initial round of scoring, another meeting was held to discuss and refine the chosen criteria. One of the modifications made during this meeting included eliminating the category of classroom setting (i.e. limitations due to classroom setup or situational factors). It was decided that these limitations would be evident from the other categories, and that all strategies could ultimately be utilized in any classroom setting. The category of 'frequency of use' was added, and it was decided that the criteria could also be grouped into the three distinct categories of 'pre-class prep,' 'in class actions,' and 'task characteristics.'

Following the review and revision of our scoring criteria, each author then preceded to complete the full written review of their two assigned strategies. Using these written reviews as a guide, two authors worked together to draft defined scoring levels for each of the scoring criteria. Each criterion was broken down into three levels of scoring, and it was from here that an initial utility score was determined for each of the strategies. Another meeting was held with all authors to discuss this initial scoring and make any needed revisions to the rubric. Subsequently, each co-author independently scored all eleven strategies according to each item on the rubric. We calculated inter-rater reliability using Fleiss' kappa values (Fleiss 1971) for multiple raters. Each category ranged from 0.611 to 0.866 indicating substantial to excellent agreement between raters (Landis and Koch, 1977). While the group of coauthors essentially views the strategies similarly, we do not propose that this rubric is universally applicable and anticipate that others may add or subtract categories or choose to give more weight to some categories.

It was decided that the criterion of research validation was of central importance to the idea of active-learning, and strategies with robust validation should be weighed more heavily in

their scoring. Consequently, the scoring for this criterion was expanded to a six-point scale, and more detailed scoring levels were drafted that included both the quality and quantity of evidence supporting a specific strategy. This was refined into a stand-alone rubric (Table 1) to provide descriptions for the six categories. Individual articles about each strategy that were cited in this paper were examined using a modified version of the GER Strength of Evidence pyramid (St. John and McNeal, this issue; see in Table S1). Modifications were made from the analysis of literature on training graduate teaching assistants by Bitting et al. (this issue). We considered 64 papers representing peer-reviewed articles that focused on one active learning strategy addressed in this article, rather than a suite of strategies together. Each article included a description of a teaching activity with either predicted or measured improvement of student success. Papers included in the references but not reviewed for strength of evidence addressed suites of active learning strategies together (e.g. McConnell et al., 2003), covered a different teaching strategy not discussed here (e.g. undergraduate research experiences, Russell et al., 2007), or were otherwise not a description of an active learning strategy (e.g. Manduca et al., 2017).

Literature supporting the use of these active learning strategies falls across all categories within the Strength of Evidence pyramid (Table S1). Two of the authors co-coded 11 articles using Table S1 and achieved 100% agreement. Subsequently, each of the remaining 53 articles were coded by one of these authors. A third of the papers (21 papers; 32.8%) were coded as examples of either practitioner wisdom or expert opinion that described a teaching strategy without an associated assessment. The next largest group of papers were represented by case studies (19 papers; 29.7%) where the author(s) analyzed data related to student learning associated with the application of a specific teaching strategy. An additional 13 papers (20.3%) were categorized as syntheses or meta-analyses assessing the impact of a strategy when used by

multiple instructors. Cohort studies (8 papers; 12.5%), representing multiple iterations of a strategy by the same instructor, and systematic reviews (3 papers; 4.7%) were less common. Unlike the teaching assistant training literature (Bitting et al., this issue), much of the research done on active learning strategies included comparison groups (19 papers; 29.7%).

Not every active learning strategy has been isolated and studied at the same level of detail. For example, gallery walks and think-pair-share were each isolated in one paper (Francek, 2006; Fitzgerald 2013), while we reviewed results from the application of the jigsaw technique and concept maps in nine and thirteen papers respectively. To provide a quick overview for practitioners, we use the efficacy levels in Table 1. These serve as a holistic ranking of the research support for a strategy as a combination of the strength of evidence available in the peer-reviewed literature and the direction of that evidence (positive, negative or neutral). For example, Slish's 2005 study of the jigsaw technique was a case study with a comparison group (2C, Table S1), but he concluded that neither the control nor experimental group was superior. Taken in conjunction with the other literature available for the jigsaw technique, the strategy overall receives a 3. In order to ensure research validation scores could be given consistently, two co-authors reviewed the articles associated with each active learning strategy on the basis of their strength of evidence rating and for the direction of evidence. They then applied the rubric shown in Table 1 to assign a research validation score for each strategy (see Table 3).

Table S1: Rubric developed and used to characterize active learning strategy papers. "General Group" headings are based on the GER Strength of Evidence pyramid (St. John and McNeal, this issue) from which Categories were determined as papers were reviewed, similar to the method described in Bitting et al. (this issue).

General Group	General Description	Category
1. Practitioner Wisdom/ Expert Opinion (teaching strategy description, with or without assessment)	A description of a use, multiple uses, or multiple iterations of a use to describe how a teaching strategy works. "Use" here may refer to implementation within a course or a lab-based study.	1A:No data on effectiveness or impact
		1B:Some form of data or comments are collected and described, but not collected or reported or analyzed systematically; trends of data are not substantiated, or data is satisfaction data without coding for additional analysis.
2. Qualitative and Quantitative Case Studies (single iteration of a research study)	Analysis of a single teaching strategy by an instructor in a class or researcher in a lab with data that allows interpretation of change related to the intervention.	2A: Data is all participant satisfaction or self-report using non-validated instruments.
		2B: Data includes validated self-report instruments, codes based on established theoretical frameworks, or objective measures of impact.
		2C: Data includes validated self-report instruments and/or objective measures of impact. Study includes comparison group of some type. Comparison group may be a separate class.
3. Qualitative and Quantitative Cohort Studies (repeated research study)	Synthesis of multiple iterations of a single teaching strategy by one instructor in a class or researcher in a lab, either presented separately or as an aggregate dataset, with data that allows interpretation of change related to the intervention.	3A: Data is all participant satisfaction or self-report using non-validated instruments.
		3B: Data includes validated self-report instruments, codes based on established theoretical frameworks, or objective measures of impact.
		3C: Data includes validated self-report instruments and/or objective measures of impact. Study includes comparison group of some type.
4. Syntheses and Meta-Analyses	4A. Synthesis of a group of individuals using a single teaching strategy, either presented separately or as an aggregate dataset, with data that allows for interpretation of change related to the interventions	
	4B: Meta-analysis of the data from multiple studies/publications to combine smaller data sets into a larger body of data that is synthesized in aggregate	
5. Systematic Reviews	Synthesis of the results of multiple studies/publications to draw broad conclusions of the group of studies	

Case Studies/Problem Learning Activities

- Stormy Weather, created by Alan Price, U. of Wisconsin-Washington County, available from the PBL Clearinghouse (http://www1.udel.edu/pblc/index.html), retrieved January 16, 2017.
- The Slippery Slope of Litigating Geologic Hazards, created by David Ozsvath, U. of Wisconsin-Stevens Point, available from the National Center for Case Study Teaching in Science (http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case_id=385&id=385), retrieved January 16, 2017.
- The National Center for Case Study Teaching in Science (NCCSTS; http://sciencecases.lib.buffalo.edu/cs/collection/). Visitors can access an overview of the case and download case materials but a paid subscription is required to review teaching notes, an answer key and comments about the cases.
- The PBL Clearinghouse at the University of Delaware (http://www1.udel.edu/inst/). This site provides both handouts for the problem itself and extensive instructor guides including learning objectives, links and citations for necessary resources and teaching notes.
- Investigative Cases (see http://bioquest.org/icbl/). Few of these cases include reference to the earth system in more than a tangential role.
- A search of NAGT's Teach the Earth site (http://serc.carleton.edu/teachearth/index.html) reveals examples of case studies and problems created for a variety of courses. While many of these examples target upper level courses, there are some exemplary examples for introductory courses, for example:
    - Earthquake Case Study, Kaatje Kraft (http://serc.carleton.edu/NAGTWorkshops/intro/activities/23588.html)
    - Hotspot Lesson, Jamie Russell, (http://serc.carleton.edu/sp/erese/final-hotspot-project.html)

Concept Maps

- The FLAG (Field-tested learning assessment guide) site has a series of pages (http://www.flaguide.org/cat/conmap/conmap1.php) developed by Michael Zeilik that explain the use of concept maps as a classroom assessment technique.
- The Florida Institute for Human & Machine Cognition (IHMC) has created free concept map tools (http://cmap.ihmc.us/) that can be downloaded for personal computers or ipads. Their site contains numerous short documents, videos and other resources designed to introduce viewers to learning with concept maps (http://cmap.ihmc.us/docs/learn.php).
    - See also *The theory underlying concept maps and how to construct and use them* (http://cmap.ihmc.us/docs/theory-of-concept-maps) by Novak and Canas (2008).
- NAGT's Teach the Earth site (http://serc.carleton.edu/teachearth/index.html) has some examples of concept maps created for a variety of courses. For example:

  - Developing concept maps (http://serc.carleton.edu/introgeo/assessment/conceptmaps.html) provides a brief introduction to concept maps that includes links to papers about their use.
  - Assessment using concept mapping (http://serc.carleton.edu/NAGTWorkshops/assess/conceptmaps.html)

Concept Sketches

- Teaching the Earth hosts a series of webinars on Effective Strategies for Undergraduate Geoscience Teaching that featured one on *Teaching and assessing in-depth understanding of fundamental concepts using concept sketches*. The 77 minute screencast of the event lead by Stephen Reynolds and Julia Johnson is available here: (http://serc.carleton.edu/NAGTWorkshops/careerdev/AcademicCareerTeach2013/march.html).

Gallery Walks

- Examples of a variety of geoscience-themed gallery walks are available on the Starting Point – Teaching Entry Level Geoscience site (http://serc.carleton.edu/introgeo/gallerywalk/index.html).
- Leah Joseph, Ursinus College, *Abandoned Mine Lands and Superfund/National Priorities List* homework assignment and related gallery walk (http://serc.carleton.edu/integrate/teaching_materials/mineral_resources/unit3pre.html).
- Gross, D., Gilbert, L., and Kreutz, K., 2016, Unit 2, Picturing Complexity in InTeGrate Systems Thinking module, retrieved, Jan 21, 2017 from http://serc.carleton.edu/163004.

Jigsaw

- Barbara Tewksbury created a great site that describes how and why to use jigsaws on the SERC Pedagogy in Action site (http://serc.carleton.edu/sp/library/jigsaws/index.html).
- The Jigsaw classroom (https://www.jigsaw.org/) discusses the general application of this teaching strategy and its development by Elliot Aronson.
- Students work together to compare examples from three countries in the *Women and Water* unit (http://serc.carleton.edu/integrate/teaching_materials/freshwater/unit4.html) of InTeGrate's Environmental Justice and Freshwater Resources module
- Students work together to compare mining activities in the *Mining, Society and Decision Making* unit (http://serc.carleton.edu/integrate/teaching_materials/mineral_resources/unit6.html) of InTeGrate's Humans' Dependence on Mineral Resources module

Minute Paper

- A description of the minute paper process by Meg Steepey, Earlham College, is provided on the Teach the Earth portal (http://serc.carleton.edu/NAGTWorkshops/assess/activities/streepey.html)

- The FLAG (Field-tested learning assessment guide) site has a series of pages (http://www.flaguide.org/cat/minutepapers/minutepapers1.php) developed by Michael Zeilik that explain the use of the minute paper as a classroom assessment technique.

Peer Instruction (Conceptests)

- Using learning assistants to support peer instruction with classroom response systems ("clickers") is available on the SERC Pedagogy in Action site (http://serc.carleton.edu/sp/library/learning_assistants/examples/example1.html).
- A discussion of conceptests and how to use them is available in the teaching methods section of the Teach the Earth site (http://serc.carleton.edu/NAGTWorkshops/teaching_methods/conceptests/index.html) including access to more than 300 conceptests, most originally developed to support an earth science class.
- Merlot's ELIXR project (http://pachyderm.cdl.edu/elixr-stories/serc-geology/) also includes short videos and classroom footage of the use of conceptests in a large earth science class at the University of Akron.
- Geoscience Concept Inventory wiki provides access to multiple choice questions on a range of geoscience topics (https://geoscienceconceptinventory.wikispaces.com/).

Role Playing

- The SERC site has examples of role playing used in mock trial scenarios (http://serc.carleton.edu/introgeo/roleplaying/examples/toxictrl.html) or in debate settings (http://serc.carleton.edu/introgeo/roleplaying/anastasio.html).
- How to design an effective role playing exercise is discussed on the SERC Pedagogy in Action site (http://serc.carleton.edu/sp/library/roleplaying/howto.html).
- More than forty examples of role playing scenarios, some embedded in larger lessons, are available on the SERC Pedagogy in Action site (http://serc.carleton.edu/sp/library/roleplaying/scenario.html).

Teaching with Models

- Interactive lecture demonstrations information available from the SERC Pedagogy in Action site at http://serc.carleton.edu/sp/library/demonstrations/index.html.
- Teaching with Models from the SERC Pedagogy in Action site at http://serc.carleton.edu/sp/library/models/index.html focuses on conceptual and mathematical models and also provides examples of teaching with visualizations and demonstrations.

Think-Pair-Share

- Description of Think-Pair-Share, including videos with examples of the strategy, available from the SERC Pedagogy in Action site (http://serc.carleton.edu/sp/library/interactive/tpshare.html).

- Examples of think-pair-share activities can be found at the Starting Point site for Teaching entry-level geoscience (http://serc.carleton.edu/introgeo/interactive/tpshareexm.html).

Lecture Tutorials

- A description of lecture tutorials was by Karen Kortz and Jessica Smay is provided on the SERC Pedagogy in Action portal (http://serc.carleton.edu/sp/library/lecture_tutorials/index.html).
    - Examples of five geological lecture tutorials can be found here, http://serc.carleton.edu/sp/library/lecture_tutorials/examples.html.

References Cited:

Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. Psychological bulletin, 76:378-383.

Landis, J.R. and Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics, 33:159-174.