

# Supplemental Material

*CBE—Life Sciences Education*

Eddy *et al.*

Supplemental Table 1. Best practices based on literature review for the implementation of the different phases of a classroom activity: Introduction, Student Engagement, and Debriefs.

		Activity Development	Activity Phases		
			Introduction	Student Engagement	Debrief
Four PORTAAL Dimensions of Best Practice	Practice	<p><i>P1.</i> Allocate class time for practice.</p> <p><i>P2.</i> Use questions/problems that align with exams.</p> <p><i>P3.</i> Include questions requiring the use of prior knowledge.</p>			<p><i>P4.</i> Have students explain their answers in front of whole class so instructor/students can provide immediate respectful feedback.</p>
	Logic Development	<p><i>L1.</i> Use questions/problems that require critical thinking.</p>	<p><i>L2.</i> Remind students to explain their answers.</p>	<p><i>L3.</i> Give students explicit time to answer alone before any other form of engagement.</p> <p><i>L4.</i> Have students discuss in groups before whole class discussion/debrief.</p> <p><i>L5.</i> Do not reveal clicker histogram or provide hints between iterations of engagement.</p>	<p><i>L6.</i> Call on students to explain their logic.</p> <p><i>L7.</i> Be sure the logic behind the right answer is explained.</p> <p><i>L8.</i> Frequently explain the why the alternative answers are wrong.</p>
	Accountability	<p><i>A1.</i> Make activity participation worth course points (ideally for participation not correctness).</p>		<p><i>A2.</i> Avoid relying on volunteers instead employ small groups work</p> <p><i>A3.</i> For whole class discussions, call on many students via cold or random call.</p>	<p><i>A3.</i> For whole class discussions, call on many students via cold or random call.</p>
	Reducing Apprehension		<p><i>R5.</i> Remind students that mistakes are a necessary part of the learning process.</p>	<p><i>R1.</i> Use random call to spread participation in discussions.</p> <p><i>R2/R6.</i> Praise class effort.</p> <p><i>R3.</i> Make sure students know their contributions to large class discussions are</p>	<p><i>R2/R3.</i> Make sure students know their contributions to debriefs are appreciated.</p> <p><i>R5.</i> Emphasize how wrong answers contribute to discussion and are valued.</p> <p><i>R6.</i> Praise students for hard</p>

				appreciated. R4. Avoid demeaning student answers.	work and effort rather than correctness/intelligence.
--	--	--	--	--	---

Supplemental Table 2. Fidelity of implementation is low across dimensions of best practice for active learning. Distribution of 25 instructors in introductory biology for each element of best practice from the research literature. Each quartile represents 25% of the data. The median is the upper bound of the 2<sup>nd</sup> quartile.

Dimension/Element	Validity (DBER researcher agreement)	Reliability (ICC)	1 <sup>st</sup> Quartile	2 <sup>nd</sup> Quartile	3 <sup>rd</sup> Quartile	4 <sup>th</sup> Quartile
<b>Dimension 1: Practice</b>						
P1. Practice Testing (Minutes Students Working)	6/7	0.993	0.9 – 3.1 min	> 3.1 – 5.7 min	> 5.7 – 12.3 min	> 12.3 – 30.8 min
P2. Alignment of Practice (Practice aligned with exams)	7/7	NA	NA			
P3. Distributed practice (% Questions cuing prior knowledge)	6/7	0.84	0.0% - 0.4%	> 0.4% - 7.5%	> 7.5% - 9.5%	> 9.5% - 21.4%
P4. Immediacy of Feedback (% Activities students provide explanations for answers to the instructor)	7/7	0.82	0.0% -33.3%	>33.3% – 60%	> 60% - 84.0%	> 84.0% - 100%
<b>Dimension 2: Logic Development</b>						
L1. % Activities focused on higher order cognitive skills	7/7	0.89	0.0% - 6.2%	> 6.2% - 17.2%	> 17.2% - 27.6%	> 27.6% - 64.1%
L2. % Activities students reminded to use logic in framing of activity	6/7	0.938	0.0% - 0.0%	0.0% - 0.0%	0.0% - 0.0%	0.0% - 0.0%
L3. % Activities students thought alone first	7/7	0.996	0.0% - 0.8%	> 0.8% - 8.3%	> 8.3% - 27.4%	> 27.4% - 82.1%
L4. % Activities students explain their logic to each other (small group work)	7/7	0.989	0.0% - 12.3%	> 12.3% - 34.2%	> 34.2% - 55.4%	> 55.4% - 75.0%
L5. % Activities answer indicated between iterations	5/7	0.932	0.0% - 0.0%	> 0.0% - 3.9%	> 3.9% - 41.7%	> 41.7% - 100%
L6. % Debriefs students provide explanations	6/7		See P3			

L7. % Activities were correct answer explained	7/7	0.88	18.7% - 70.2%	> 70.2% - 75.2%	> 75.2% - 85.9%	> 85.9% - 93.4%
L8. % Activities were alternative answers discussed	7/7	0.98	0.0% - 1.0%	> 1.0% - 10.2%	> 10.2% - 13.3%	> 13.3% - 29.8%
<b>Dimension 3: Accountability</b>						
A1. % Activities worth course points	7/7	0.993	0.0% - 0.3%	> 0.3% - 26.8%	> 26.8% - 49.0%	> 49.0% - 90.5%
A2. % Activities with some kind of accountability (not volunteer)	7/7	See L4				
A3. % Whole class discussions Cold/Random Call	6/7	0.852	0.0% - 0.0%	0.0% - 0.6%	0.0% - 0.6%	> 0.6% - 56.9%
<b>Dimension 4: Apprehension Reduction</b>						
R1. % Activities with Random Call	6/7	See A3				
R2. % debriefs class received positive feedback and/or encouragement	6/7	0.863	0.0% - 0.0%	0.0% - 7.5%	7.5% - 9.5%	9.5% - 21.4%
R3. % student responses with positive feedback and/or encouragement	6/7	0.86	0.0% - 0.0%	0.0% - 12.2%	> 12.2% - 31.7%	> 31.7% - 100%
R4. % student responses with negative feedback	6/7	NA**	0.0% - 0.0%	0.0% - 0.0%	0.0% - 0.0%	0.0% - 0.0%
R5. Emphasize errors natural/instructional (error framing)	7/7	NA	NA			
R6. Emphasizing hard work vs. ability	7/7	NA	NA			

## Validation: BER reviewer comments

Although comments were collected on all elements, only one element was flagged for misalignment between element and observation by 2 of the 7 reviewers. This element was L5: Hints or Histograms given between iterations of student engagement in the activity. Both Reviewers pointed out the value of potentially eliminating 2 or more distractors to have students focus on a single contrast, especially if there is a clear divide in the class between two responses. We have not seen a paper addressing this specific condition. We recognize this practice has the potential to be powerful when done occasionally. The papers we encountered pointed out that when an instructor frequently eliminates options, students learn that they do not have to work as hard at answering the question. Rare events may be fine. In addition, if students discuss the problems in small groups after the initial clicker questions, and most of the class is split between two answers, then most of the groups will likely naturally gravitate towards discussing those two option (as most of the people in the group would have chosen one or the other of them) without instructions to do that from the instructor. So, instructor elimination may not be necessary in this case.

The comments collected on the other elements came from both dissenters and those who agreed with the element. Reviewing all the comments, we found that most of them fell into four categories. The first category consisted of statements indicating interest in qualifying the element to make it even better by capturing measures of quality. Many of these qualifying statements were addressed by other elements, but some would require observers to have a deeper knowledge of the course (what has come before or what is coming after, course goals) and pedagogical content knowledge to assess alignment between class practices and those goals. These are the type of measures the RTOP tool excels at capturing, but are outside the bounds of PORTAAL. PORTAAL is intended for users without deep pedagogical content knowledge. Thus, these points were considered and we explored whether we thought observers could objectively evaluate the criteria. If we could not come up with a way to do this, we did not add the qualifications to the tool.

A second set of comments focused on having the tool use information from events occurring outside of class (such as homework assignments or participation online). Although these suggestions were appreciated and were likely pointing out important considerations, they were outside the scope of PORTAAL which is only intended to explore in class practices.

The third type of comments questioned the rate at which different elements need to be present in the classroom. We whole heartedly agree with these comments. Currently, we do not know how frequently certain elements need to be present in the classroom (for example: how many clicker questions need to be higher order to help develop logic? How often do instructors need to praise to reduce apprehension?). We have tried to make this clear in the body of the paper and believe this is an area for future work that hopefully PORTAAL can help contribute to.

The final category was composed of comments based in the reviewer's own personal experience. These comments we did not attempt to address as one of the aims of PORTAAL was only to use evidence from peer reviewed literature.