Don't have a pre-post test, use a Bloom's rubric! The development and validation of a rubric for "blooming" assessments to measure student learning

Janet Casagrand Ph.D. and Kate Semsar Ph.D.

Integrative Physiology, University of Colorado at Boulder

SEI The Science Education Initiative at the University of Colorado

Purpose and Research Question

Purpose:

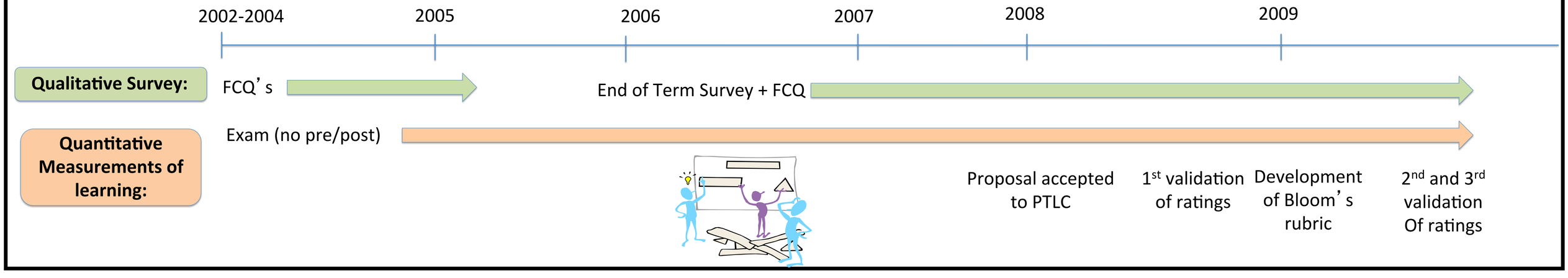
- To compare student performance on assessments from different semesters when a concept or pre-post survey does not exist or when other assessment tools are not appropriate

Proposal submitted and accepted to the President's Teaching and Learning Collaborative (PTLC)

Methods

- Independent, blinded raters (n=3)
- 3 rounds of validating Bloom's level of course questions (n=155): round 1 with standard rubric, round 2 with preliminary flow chart rubric, round 3 with finalized flow chart rubric
- Inter-rater reliability and percent agreement between raters scored at each round

Development of Bloom's rubric



Bloom's rubric validation & example use

Table 1. Percent agreement between raters

	without flow rubric	with flow rubric	final rating with revised flow rubric
Total number of questions	26	26	155
Percent > 2 raters agree	92%	96%	93.5%
Percent 3 raters agree	31%	54%	53.5%
No consensus	8%	4%	6.5%

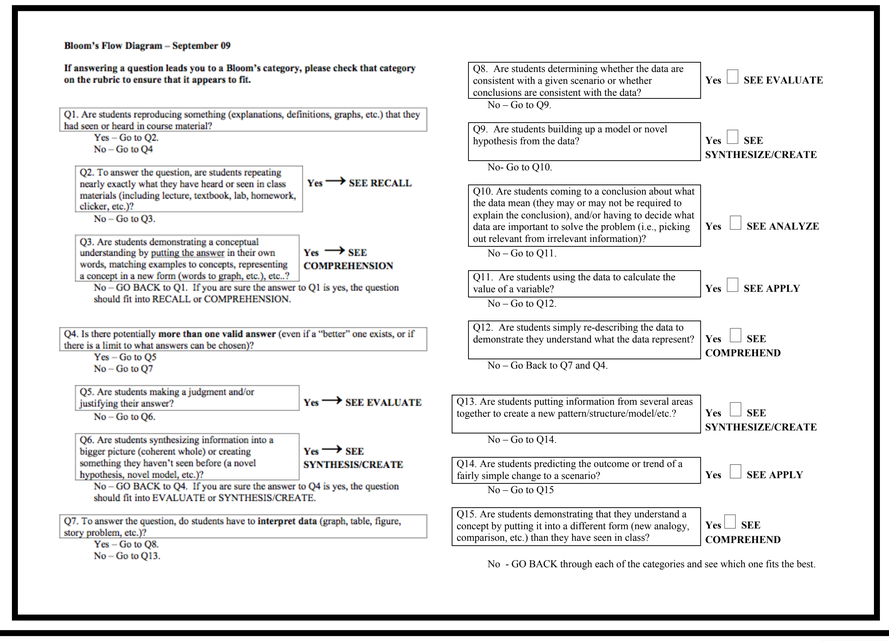
Table 2. Percent agreement between a rater and average question rating

	without flow rubric	with flow rubric	final rating with revised flow rubric
Number of questions	24 (of 26)	25 (of 26)	145 (of 155)
Rater 1	63%	84%	83%
Rater 2	83%	84%	92%
Rater 3	88%	88%	86%
Average of raters	78%	85%	87%

Example "Blooming" of exam question

What are the directions of the chemical, electrical, and net driving forces acting on K+ when the membrane potential is -55mV?

Bloom's Rubric Q1: *No* → Q4
Bloom's Rubric Q4: *No* → Q7
Bloom's Rubric Q7: *No* → Q13
Bloom's Rubric Q13: *No* → Q14
Bloom's Rubric Q14: *Yes* → Apply



Bloom's Flow Diagram – September 09

If answering a question leads you to a Bloom's category, please check that category on the rubric to ensure that it appears to fit.

Q1. Are students reproducing something (explanations, definitions, graphs, etc.) that they had seen or heard in course material?
Yes – Go to Q2.
No – Go to Q4

Q2. To answer the question, are students repeating nearly exactly what they have heard or seen in class materials (including lecture, textbook, lab, homework, clicker, etc.)? **Yes → SEE RECALL**
No – Go to Q3.

Q3. Are students demonstrating a conceptual understanding by putting the answer in their own words, matching examples to concepts, representing a concept in a new form (words to graph, etc.), etc..? **Yes → SEE COMPREHENSION**
No – GO BACK to Q1. If you are sure the answer to Q1 is yes, the question should fit into RECALL or COMPREHENSION.

Q4. Is there potentially **more than one valid answer** (even if a "better" one exists, or if there is a limit to what answers can be chosen)?
Yes – Go to Q5
No – Go to Q7

Q5. Are students making a judgment and/or justifying their answer? **Yes → SEE EVALUATE**
No – Go to Q6.

Q6. Are students synthesizing information into a bigger picture (coherent whole) or creating something they haven't seen before (a novel hypothesis, novel model, etc.)? **Yes → SEE SYNTHESIS/CREATE**
No – GO BACK to Q4. If you are sure the answer to Q4 is yes, the question should fit into EVALUATE or SYNTHESIS/CREATE.

Q7. To answer the question, do students have to **interpret data** (graph, table, figure, story problem, etc.)?
Yes – Go to Q8.
No – Go to Q13.

Q8. Are students determining whether the data are consistent with a given scenario or whether conclusions are consistent with the data? **Yes ☐ SEE EVALUATE**
No – Go to Q9.

Q9. Are students building up a model or novel hypothesis from the data? **Yes ☐ SEE SYNTHESIZE/CREATE**
No- Go to Q10.

Q10. Are students coming to a conclusion about what the data mean (they may or may not be required to explain the conclusion), and/or having to decide what data are important to solve the problem (i.e., picking out relevant from irrelevant information)? **Yes ☐ SEE ANALYZE**
No – Go to Q11.

Q11. Are students using the data to calculate the value of a variable? **Yes ☐ SEE APPLY**
No – Go to Q12.

Q12. Are students simply re-describing the data to demonstrate they understand what the data represent? **Yes ☐ SEE COMPREHEND**
No – Go Back to Q7 and Q4.

Q13. Are students putting information from several areas together to create a new pattern/structure/model/etc.? **Yes ☐ SEE SYNTHESIZE/CREATE**
No – Go to Q14.

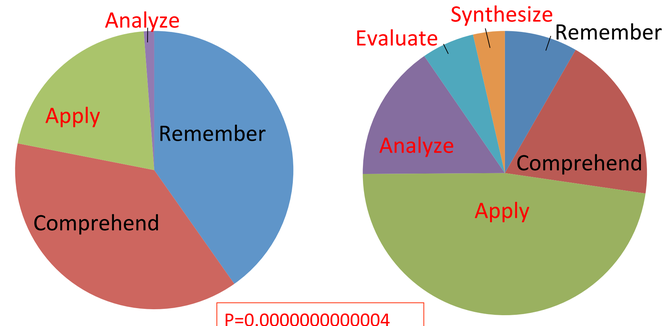
Q14. Are students predicting the outcome or trend of a fairly simple change to a scenario? **Yes ☐ SEE APPLY**
No – Go to Q15

Q15. Are students demonstrating that they understand a concept by putting it into a different form (new analogy, comparison, etc.) than they have seen in class? **Yes ☐ SEE COMPREHEND**

No - GO BACK through each of the categories and see which one fits the best.

Application of rubric

Bloom's Analysis of IPHY 4720 Exam Questions



P=0.0000000000004

Fall 2004
78% remember & comprehend
22% apply & analyze

Fall 2008
27% remember & comprehend
64% apply & analyze
9% evaluate & synthesize

Assessing a moving target? Measure how target changes!

Table 3. Exam scores before/after reform. Of the 84 questions on the three exams, 72 questions were different.

	before reform (F04)	after reform (F08)
Exam 1	72.9 ± 11.7* (n=80)	66.7 ± 16.3* (n=97)
Exam 2	74.8 ± 14.1 (n=80)	74.2 ± 13.2 (n=97)
Exam 3	70.1 ± 12.3 (n=79)	70 ± 14.2 (n=97)

As exams kept changing to maintain a consistent average (Table 3), learning could not be assessed directly. Therefore, instead of using exam scores to measure learning, we measured the changes in the exam itself (see Bloom's analysis of IPHY 4720 Exams above.) As students were able to maintain consistent scores on a more challenging assessment, we can indirectly state that IPHY 4720 reform led to increases in student learning and performance.

Conclusions

1. Between independent raters, use of Bloom's rubric is reliable.
2. A validated Bloom's rubric can be used to compare student performance on assessments when other assessment tools are inappropriate or unavailable.