Can the effectiveness of teaching methods be measured with final exam scores?

Georg Rieger

Physics & Astronomy, UBC

What is measured?

- ‘Performance’ score on final exam:

$$performance = \frac{exam\ percentage\ \times Bloom's\ level}{average\ Bloom's\ level\ (2.87)}$$

- Phys 100 (2006 – 2013)
- N = (640 – 840) students
- Style of final exam has not changed since 2006.

Fig. 1: Average final exam percentage and average final exam performance. Error bars reflect the standard deviation of the 2010 – 2014 data.

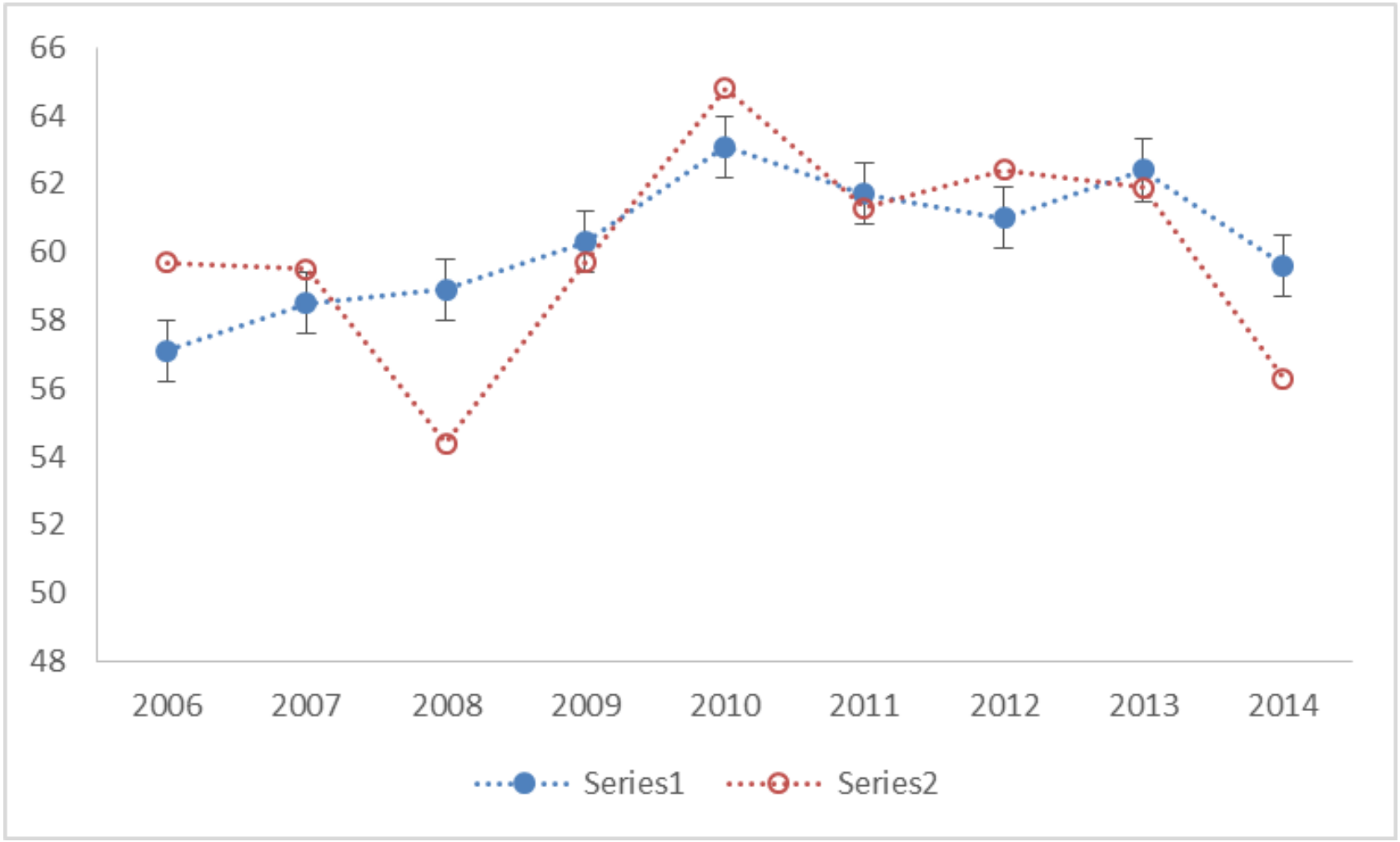

Table 1. Format of the final exams in Physics 100 and average scores. The number of multiple-choice N (MC) questions is shown in column2; the number of parts in problem questions N (PQ) is shown in column3. Columns 4 and 5 show the percentage weight of multiple-choice (MC %) and problem questions (PQ %) contributed to the final exam scores, respectively. The average exam score is in column 6 and the average Bloom's level of each final exam is shown in column 7. The corresponding exam performance score = (Bloom average x Exam average)/(Average Bloom's level) is shown in column 8.

Year	N (MC)	N (PQ)	MC %	PQ %	Exam Average %	Bloom's Level	Performance % = col6*col7/2.87
2006	10	15	40	60	59.7	2.75	57.2
2007	10	11	38	62	59.5	2.82	58.5
2008	9	10	47	53	54.4	3.11	58.9
2009	10	11	28	72	59.7	2.90	60.3
2010	16	9	50	50	64.8	2.80	63.2
2011	16	16	46	54	61.3	2.88	61.5
2012	15	16	38	62	62.4	2.81	61.1
2013	22	14	48	52	61.9	2.89	62.3
2014					56.3	3.04	59.6

Bloom's Levels

- Evaluated by single rater (me)

- Two sources:

- Bloom's level chart with action words (from Carl's learning goal presentation)
- Blooming tool (Casagrand and Semsar, U of Colorado, unpublished)

Table 2. Column 2 shows the re-normalized performance = performance/(average Bloom's level)*100. Columns 3 and 4 show CLASS results for pre-/posts shift in the general problem solving category and the overall shift, both for the favorable category. Column 5 shows the overall CLASS score (fav.) at the end of a term. The last column shows the new pedagogies introduced into the course. All new pedagogies are still in use. For example open-book exams are used since 2006. (Clickers and peer-instructions were introduced in 2002.) The CLASS data in columns 3 – 5 is corrected for the average grade dependence. {The result of the correction is shown in brackets.}

Year	Normalized Performance	CLASS-PS_Shift (fav.) {adjusted}	CLASS-All_Shift (fav.) {adjusted}	CLASS-ALL_Post (fav.) {adjusted}	New Pedagogy
2006	57.1	-5.5 ± 2.9 {-8.1}*	-2.7 ± 1.7 {-4.2}*	45.7 ± 2.0 {42.4}*	Open book midterm and final exams *Small sample (N=91); CLASS grade average very different from Course grade average (- 7.8)
2007	58.5	0.5 ± 1.1 {-0.3}	-2.5 ± 0.7 {-2.7}	51.0 ± 0.9 {50.0}	Context-rich tutorials and group work; Learning Goals
2008	58.9	0.8 ± 1.2 {0.5}	-2.0 ± 0.8 {-2.2}	47.7 ± 0.9 {47.2}	Custom textbook
2009	60.3	-2.7 ± 1.2 {-3.0}	-5.4 ± 0.7 {-5.6}	47.4 ± 0.9 {47.0}	Pre-class reading assignments
2010	63.1	4.1 ± 1.4 {3.4}	-0.9 ± 0.9 {-1.4}	51.0 ± 1.1 {49.4}	Worksheets in lecture
2011	61.7	4.2 ± 1.1 {3.7}	0.5 ± 0.7 {0.1}	52.5 ± 0.9 {50.4}	
2012	61.0	3.7 ± 1.4 {2.5}	-0.3 ± 0.8 {-0.9}	54.5 ± 1.0 {53.1}	Two-stage midterm exams
2013	62.4	No data	No data	No data	

Analysis 2:

Another way to compare the data is to simply compare the averages and standard deviations for the (2006 – 2009) and (2010 – 2013) periods, which correspond to the years before and after introducing worksheets into the lecture portion. Table 3 shows the results.

Table 3. Average exam scores and performance scores aggregated for two time periods.

Period	Exam Score	STD DEV	Perfor-mance	STD DEV
2006 – 2009	58.3%	2.6%	58.7%	1.3%
2010 – 2013	62.6%	1.5%	62.1%	0.9%

Carl's Bloom's Level Chart

(Learning Goals workshop, UBC PHAS, May 2007)

Bloom's Taxonomy of the Cognitive Domain
(~= content+skills+habits of mind)

1. Factual Knowledge: remember and recall factual information
   Define, List, State, Label, Name, Describe
2. Comprehension: demonstrate understanding of ideas, concepts
   Describe, Explain, Summarize, Interpret, Illustrate
3. Application: apply comprehension to unfamiliar situations
   Apply, Demonstrate, Use, Compute, Solve, Predict, Construct, Modify
4. Analysis: break down concepts into parts
   Compare, Contrast, Categorize, Distinguish, Identify, Infer
5. Synthesis: transform, combine ideas to create something new
   Develop, Create, Propose, Formulate, Design, Invent
6. Evaluation: think critically about and defend a position
   Judge, Appraise, Recommend, Justify, Defend, Criticize, Evaluate

Higher level: Require deeper conceptual understanding

Table 2. Bloom's Dichotomous Key (BDK). (Casagrand and Semsar, U of Colorado)	
• Categorize the question based on what students are being asked to do, not on how challenging the question may be. (For example, a 'comprehend' question for a difficult concept could be a more challenging problem than an 'analyze' question on an easier concept.) • Evaluate questions with reference to what material we know students were exposed.	
Q1. Could students memorize the answer to this specific question?	
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 <u>putting the answer</u> 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 solution* (even if a "better" one exists, or if there is a limit to what solutions can be chosen)?	
Yes – Go to Q5. No – Go to Q8.	
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 SYNTHESIZE/CREATE
No – Go to Q7.	
Q7. Are students being asked to	Yes → SEE ANALYZE

Q8. To answer the question, do students have to interpret data (graph, table, figure, story problem, etc.)?	
Yes – Go to Q9. No – Go to Q14.	
Q9. 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 Q10.	
Q10. Are students building up a model or novel hypothesis from the data?	Yes → SEE SYNTHESIZE/CREATE
No- Go to Q11.	
Q11. 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 Q12.	
Q12. Are students using the data to calculate the value of a variable?	Yes → SEE APPLY
No – Go to Q13.	
Q13. Are students re-describing the data to demonstrate they understand what the data represent?	Yes → SEE COMPREHEND
No – Go Back to Q8 and Q4.	
Q14. Are students putting information from several areas together to create a new pattern/structure/model/etc.?	Yes → SEE SYNTHESIZE/CREATE
No – Go to Q15.	
Q15. Are students predicting the outcome or trend of a fairly simple change to a scenario?	Yes → SEE APPLY
No – Go to Q16.	
Q16. Are students demonstrating that they understand a concept by putting it into a different form (new example, analogy, comparison, etc.) than they have seen in class?	Yes → SEE COMPREHEND
No - GO BACK through each category or refer category descriptions to see which fits the	